## 

Desafío de Tripulaciones

13/03/2024

**─**

Randbee Consultants

http://randbee.com

@RandbeeCo

[“NOWCASTING”: haciendo predicciones sobre el tiempo presente 2](#_Toc161320534)

[Antecedentes sobre el proyecto 2](#_Toc161320535)

[Los enfoques tradicionales de “forecasting” usan estadísticas oficiales como indicadores precisos de la actividad social y económica. Sin embargo, muchas de estas estadísticas oficiales sólo están disponibles en frecuencias mensuales, trimestrales o incluso anuales. 2](#_Toc161320536)

[Esto significa que pueden ser insuficientes para detectar cambios inesperados durante periodos de gran incertidumbre (e.g., pandemia de COVID-19, guerra de Ucrania). 2](#_Toc161320537)

[Contexto: el desafío de Randbee Consultants 2](#_Toc161320538)

[Cuál es nuestra propuesta de trabajo para el “Desafío de tripulaciones” 2](#_Toc161320539)

[Principales tareas a realizar por los estudiantes 2](#_Toc161320540)

[Recursos útiles para el desarrollo de las tareas 3](#_Toc161320541)

[Output variable o variable respuesta: integración en el mercado laboral 3](#_Toc161320542)

[Otros recursos útiles 3](#_Toc161320543)

[Extensión de la serie temporal 4](#_Toc161320544)

[**Consideraciones estadísticas** 4](#_Toc161320545)

[**Cómo pueden extraerse los datos de Google Trends** 5](#_Toc161320546)

[**Cómo pueden descargarse los datos de GDELT** 6](#_Toc161320547)

[**Otros recursos** 7](#_Toc161320548)

[**Palabras clave y temas** 8](#_Toc161320549)

[**Otros recursos** 9](#_Toc161320550)

[**Recomendaciones sobre el diseño del modelo** 10](#_Toc161320551)

[LSTM 11](#_Toc161320552)

[**Otros recursos** 12](#_Toc161320553)

# 

# “NOWCASTING”: haciendo predicciones sobre el tiempo presente

## Antecedentes sobre el proyecto

## Los enfoques tradicionales de “forecasting” usan estadísticas oficiales como indicadores precisos de la actividad social y económica. Sin embargo, muchas de estas estadísticas oficiales sólo están disponibles en frecuencias mensuales, trimestrales o incluso anuales.

## Esto significa que pueden ser insuficientes para detectar cambios inesperados durante periodos de gran incertidumbre (e.g., pandemia de COVID-19, guerra de Ucrania).

Los modelos de “nowcasting” difieren de los modelos predictivos, puesto que, en lugar de intentar predecir el futuro, se centran en predecir el presente, “rellenando” los huecos de información existentes para así proporcionar datos a una mayor resolución temporal.

## Contexto: el desafío de Randbee Consultants

Generar modelos de “nowcasting” basados en técnicas de “machine learning” (ML): e.g., deep neural network (DNN), stacked ensembles regression (SE), random forests (RF), and extreme gradient boosting (XGB), long–short term memory (LSTM) neural network).

Estos modelos se aplicarán para predecir a tiempo real un conjunto de variables socio-económicas de entre un conjunto predefinido por la Unión Europea.

### Cuál es nuestra propuesta de trabajo para el “Desafío de tripulaciones”

Generar un modelo de “nowcasting” basado en una técnica de ML para predecir a tiempo real una de las variables socio-económicas (i.e., integración de los inmigrantes en el mercado laboral) a partir de “big data” obtenidos de “Google Trends” y de “GDELT” (‘Global Database on Events Location and Tone’).

### Principales tareas a realizar por los estudiantes

1. Implementar una API que permita extraer información sobre vólumenes de consultas de “Google search data” agregadas bajo la forma de “Google Trends” para un conjunto predefinido de “temas” y “categorías” en distintos periodos de tiempo y a determinadas frecuencias temporales.
2. Implementar una API que permita extraer información sobre “sentiment indicators” a partir de la base de datos GDELT, en forma de “Article tone” y “Topic popularity rate”.
3. Construir (y validar) un modelo de “nowcasting” usando técnicas de ML (LSTM neural network) para la variable socioeconómica señalada utilizando los datos the Google Trends y de GDELT previamente extraídos como predictores.

# Recursos útiles para el desarrollo de las tareas

### Output variable o variable respuesta: integración en el mercado laboral

La variable a modelar será “integración en el mercado laboral”, cuantificada a partir de datos proporcionados por Eurostat. En concreto, la serie temporal del indicador: *Population by sex, age, citizenship and labour status (1 000) -*[lfsa\_pganws](https://doi.org/10.2908/LFSA_PGANWS), medido como tasa de actividad por sexo; edad y país de nacimiento, que presenta una frecuencia anual. Estos datos están disponibles desde el año 1995.

Se trata de datos en formato JSON-stat que, en principio, pueden descargarse a través de la [API de Eurostat](https://wikis.ec.europa.eu/display/EUROSTATHELP/API+-+Getting+started). La API permite descargar también estos datos en otros formatos.

Existen también paquetes de R y Python que permiten descargar los datos de Eurostat a través de la API:

[eurostat R package](https://ropengov.github.io/eurostat/articles/eurostat_tutorial.html); <https://github.com/eurostat/restatapi>

[eurostat 1.0.4](https://pypi.org/project/eurostat/)

[eurostatapiclient 0.3.0](https://pypi.org/project/eurostatapiclient/)

A pesar de que los datos aparecen desagregados a varios niveles, tomaremos la tasa de actividad para ciudadanos extranjeros citizen [FOR], considerando toda la población (la suma de todas las clases de edad) y sólo la población activa wstatus [ACT] para cada país de residencia geo [todas las categorías] y para cada clase de sexo [F, M] por separado.

### Otros recursos útiles

<https://www.kaggle.com/code/crischir/eurostat-api>

<https://towardsdatascience.com/using-eurostat-statistical-data-on-europe-with-python-2d77c9b7b02b>

https://github.com/eurostat/PRost

# Extensión de la serie temporal

La serie temporal a analizar y que se puede utilizar para calibrar y validar los modelos estará limitada por la longitud de la serie temporal de datos más corta: la variable respuesta está disponible desde 1995, Google Trends presenta datos desde el año 2004 y GDELT GKG desde 2014.

**Sobre Google Trends**

### **Consideraciones estadísticas**

Google Trends no proporciona el número real de búsquedas de un tema de interés en un momento y un lugar concretos, sino que muestra el "interés a lo largo del tiempo" pues proporciona series temporales de volumen relativo de búsqueda (Search Volume Index, o SVI), normalizados para un periodo y región geográfica, utilizando una escala de 0 a 100. Un valor de 100 representa la popularidad máxima de una búsqueda en particular. De manera que, por ejemplo, un valor de 30 indica que un interés de búsqueda particular fue un 30% más popular en comparación con el pico de la actividad de búsqueda. Google también muestra, cuando es posible, las 25 búsquedas principales y los temas vinculados a cualquier tema o consulta en particular. Las consultas principales son las consultas más buscadas por los usuarios en la misma sesión en cualquier momento y lugar determinados.

En la extracción, cada punto de datos es filtrado por el rango temporal especificado (ya sea diario, semanal o mensual) y el rango geográfico (ya sea país o ISO 3166-2), y se divide por el número total de búsquedas para obtener la medida de popularidad relativa. Las cifras se basan en una distribución uniforme sobre una muestra aleatoria de búsquedas de Google actualizada una vez al día desde 2004. Por lo tanto, puede haber diferencias entre extracciones con idénticos parámetros de filtrado. De este modo, una muestra difiere de otra extraída con posterioridad utilizando el mismo conjunto de palabras clave, incluso aunque las palabras clave usadas, el intervalo de tiempo y la región estén configurados igual.

Para soslayar este problema, una posible solución es utilizar promedios de muchas muestras para mejorar la consistencia de la serie. De este modo, siguiendo a Woloszko (2020) podría hacerse una **secuencia de seis solicitudes idénticas en términos de área geográfica y periodo temporal (“location” y “timestamp”), en una secuencia de 10 minutos**. Estas seis muestras de SVIs serán posteriormente promediadas y será sobre este valor promedio con el que se construirá el modelo.

Las consultas se pueden hacer utilizando términos/palabras clave pero también por temas (“topic”) y/o por categorías (“categories”). Como las palabras clave son específicas de cada idioma, es mejor **hacer la extracción de los datos por “topic”**. Esta extracción por “topic” puede hacerse en [PyTrends](https://github.com/GeneralMills/pytrends) (ver más detalles a continuación en los métodos para extraer datos de Google Trends) utilizando como parámetro de la búsqueda el topic\_id en lugar del topic\_name:

[Por ejemplo](https://stackoverflow.com/questions/47389000/pytrends-how-to-specify-a-word-as-a-topic-instead-of-a-search-term), para el topic\_name “Holidays”, se utilizaría su topic\_id /m/03gkl:

pytrends.build\_payload(kw\_list=["/m/03gkl"], cat=0, timeframe='today 5-y', geo='', gprop='')

pytrends.interest\_over\_time()

**Posteriormente, podemos refinar la petición de datos utilizando** [categorías](https://github.com/pat310/google-trends-api/wiki/Google-Trends-Categories) específicas dentro de cada “topic” que indicamos más abajo.

El intervalo temporal al que pertenecen los datos (diario, semanal o mensual) también se puede elegir. **Usaremos un intervalo mensual**. En cuanto al aspecto espacial, los datos deberían descargarse para **todos los países** incluidos en la base de datos de la variable respuesta.

### **Cómo pueden extraerse los datos de Google Trends**

No hay una API oficial para Google Trends, aunque hay diversas herramientas no oficiales que lo que hacen es simplemente “web-scrapping” a partir de Google Trends, como, por ejemplo, la librería [PyTrends](https://github.com/GeneralMills/pytrends) de Python, que permite interaccionar con los datos de Google Trends. Este es, con diferencia, el servicio API de Google Trends más grande, más popular y mejor mantenido. A través de estas APIs es posible acceder a los datos de la Búsqueda de Google mediante el Índice de volumen de búsqueda (SVI), tanto de consultas como de temas que, como decíamos, está normalizado según el periodo y la ubicación de la consulta. Además, en Python existen otras librerías y herramientas de “web-scrapping”, como *Beautiful Soup* y *Selenium*. Estas herramientas permiten a los usuarios simular interacciones humanas con el sitio web de Google Trends, automatizando el proceso de recopilación de datos.

Para R, por ejemplo, tenemos el paquete *gtrendR*: Analice y visualice datos de la API de Google Trends for Health con R <https://gtrendr.readthedocs.io/en/latest/>

**Sobre GDELT**

La Global Database of Events, Language, and Tone (GDELT) es una plataforma abierta de *big data* de noticias recopiladas a nivel mundial, que contiene datos estructurados extraídos de fuentes de radiodifusión, impresas y web en más de 65 idiomas. GDELT recopila todas las noticias en línea publicadas cada 15 minutos, traduciéndolas automáticamente al inglés y codificando cada artículo de noticias por tema, sentimiento y tono, ubicación y entidad (organizaciones y personas), entre otros.

GDELT genera varias bases de datos, siendo la GDELT Global Knowledge Graph (GKG) la más relevante para este trabajo. La GDELT GKG es una base de datos de artículos de noticias donde [se codifican entidades, temas, ubicaciones y tono](https://hackage.haskell.org/package/gdelt-0.1.0.0/docs/GDELT-V2-GKG.html). De este modo, GDELT permite medir el volumen y tono de las noticias sobre un tema en un país. Lo que hace que el uso de GDELT sea muy atractivo, es su amplia cobertura de eventos mundiales, lo que permite analizar fenómenos en diferentes regiones del mundo, ya que la base de datos traduce noticias al inglés desde más de 65 idiomas. En concreto, la base de datos GDELT GKG recoge artículos desde el año 2014.

Para construir esta base de datos, se procesa cada artículo de noticias, del que se extrae una lista de todas las personas, organizaciones, ubicaciones y temas en el mismo, y se concatena para formar una "clave" única que representa esa combinación particular de nombres, ubicaciones y temas. Todos los artículos que contienen la misma combinación única de nombres, ubicaciones y temas, independientemente de cuán similar sea el resto del texto, se agrupan en un "conjunto de nombres" (i.e., “nameset”).

Los datos GDELT se utilizan en una amplia gama de campos de investigación, tanto como variables directas como indirectas, a veces incluso para construir ambas. GDELT GKG se ha usado, por ejemplo, para extraer el número diario de artículos periodísticos que cubren un tema determinado.

utilizaremos **GDELT GKG para, a partir de unos temas que definiremos después, filtrar noticias relacionadas con dichos temas que están relacionados con la variable repuesta**, aunque limitándonos exclusivamente a las noticias que se han publicado sobre los países europeos de interés de acuerdo con la variable respuesta.

### **Cómo pueden descargarse los datos de GDELT**

A diferencia del flujo de eventos GDELT principal, que está diseñado para poder ser importado de manera directa a los principales paquetes estadísticos como R, el formato de archivo GKG requiere de una descarga más sofisticada y de cierto preprocesamiento por lo que habrá que usar Python para extraer y reprocesar los datos antes de importarlos a un paquete estadístico. Por lo tanto, para extraer datos de GKG, los usuarios deben aplicar procesamiento de texto y “scripting”. La [GKG](http://data.gdeltproject.org/documentation/GDELT-Global_Knowledge_Graph_Codebook-V2.pdf) ha sido optimizado para un escaneo rápido, almacenando un registro por línea y usando un formato delimitado por tabulaciones para separar los campos.

Un enfoque ampliamente utilizado para analizar datos GDELT GKG directamente sin necesidad de descargarlos es utilizar la interfaz [Big Query](https://quaintitative.com/extract_gdelt_gkg_bigquery/). Sin embargo, se invita a los estudiantes a explorar otras formas de extraer datos que puedan ser más ventajosas.

Por ejemplo, mediante Elasticsearch, que es un almacén de documentos popular y eficiente construido sobre la librería de búsqueda [Apache Lucene](https://lucene.apache.org/). Esta herramienta proporciona búsquedas y análisis en tiempo real para diferentes tipos de estructuras de datos complejas, como texto, datos numéricos o datos geoespaciales, que han sido serializadas, tales como documentos JSON. Esta herramienta, permite indexar datos facilitando búsquedas rápidas de los mismos a través de una API REST simple.

Otras posibilidades son:

En Python:

* gdeltdoc: Doc API Client para GDELT 2.0. Documentaciones disponibles aquí https://github.com/alex9smith/gdelt-doc-api
* gydelt: para la Global Knowledge Graph (GKG). Documentación disponible aquí: <https://gydelt.readthedocs.io/en/latest/README.html>

R cuenta con dos librerías que permiten a los usuarios interactuar con partes de los datos de GDELT fuera de Big Query:

* gdeltr
* GDELTherramientas

Sin embargo, estos paquetes son antiguos, incompletos y difíciles de usar.

En su lugar, hay una interfaz moderna del Proyecto R GDELT:

* [gdeltr2](https://github.com/abresler/gdeltr2). Permite a los usuarios de R acceder fácilmente a los datos de GDELT, lo que permite visualizaciones y análisis de datos más rápidos.

### **Otros recursos**

<https://newsinitiative.withgoogle.com/es-es/resources/trainings/basics-of-google-trends/>

<https://lazarinastoy.com/the-ultimate-guide-to-pytrends-google-trends-api-with-python/>

<https://www.premiumleads.com/en/blog/seo/how-to-get-google-trends-data-with-pytrends-and-python/>

<https://www.topicfinder.com/methods-scrape-google-trends/>

<https://medium.com/@atakanguney94/a-brief-introduction-into-gdelt-global-database-of-events-language-and-tone-e96b0c64d03a>

<http://data.gdeltproject.org/documentation/GDELT-Event_Codebook-V2.0.pdf>

<https://github.com/maelfabien/Cassandra-GDELT-Queries>

https://hackage.haskell.org/package/gdelt-0.1.0.0/docs/GDELT-V2-GKG.html

## **Palabras clave y temas**

Para capturar tendencias en varios idiomas **podemos usar los temas (“topics/categories”) proporcionados por Google Trends**. Los temas son encabezados temáticos predefinidos que agrupan grafías alternativas, nombres en otros idiomas y palabras relacionadas con el mismo concepto bajo una sola etiqueta. Además, el uso de los “topics/categories” en lugar de “keywords” permite reducir la ambigüedad y la interpretabilidad de los términos de búsqueda individuales, evitando correlaciones espúreas con la variable respuesta.

Una lista simplificada de “topics/categories” puede estar inspirada en la dada por [Kohns and Bhattacharjee 2023](https://doi.org/10.1016/j.ijforecast.2022.05.002):

* Labour Market: topic - Unemployment benefits, topic - jobs, topic - Unemployment, Welfare & unemployment
* Credit, Loans & Personal Finance: topic - Investment, topic - Mortgage, topic - Interest rate, Credit & lending, Investing
* Jobs & Education: Education, Jobs
* Business & Industrial Activity: Construction, consulting & contracting, Business services, Transportation & logistics, manufacturing
* Health: que puede refinarse utilizando la categoría, Medical Facilities & Services

De GDELT podemos extraer indicadores sobre la atención de los medios (**número de noticias** relacionados con los temas de interés), el **tono** (utilizando tres dimensiones: positivo, negativo e incertidumbre) y posteriormente normalizando estas métricas por el número total de artículos publicados sobre un país y el número de noticias relacionadas con los temas de interés. **Seleccionaremos solo la información publicada sobre los países de interés (i.e., “location”) y nos centraremos en los artículos que tienen al menos tres palabras clave relacionadas con cada uno de los temas (“themes”) especificados a continuación**. Además, **omitiremos del análisis los artículos demasiado cortos (sólo artículos con un recuento mínimo de palabras = 500), y si es posible, se considerarán sólo artículos publicados en medios que hayan existido durante todo el periodo** de extracción, desde el inicio hasta el final. Los datos diarios que incluyan mención a los temas de interés, se **normalizarán dividiendo por el número total de artículos publicados** ese día. Estos datos diarios se promediarán después para obtener datos a una resolución mensual. Una descripción sobre la información y los campos contenidos en la GDELT GKG puede encontrarse [aquí](https://hackage.haskell.org/package/gdelt-0.1.0.0/docs/GDELT-V2-GKG.html).

La [taxonomía temática del World Bank Group](https://vocabulary.worldbank.org/taxonomy/1737.html), que consta de más de 2198 entradas, se incluye en GDELT Global Knowledge Graph 2.0. Esta taxonomía cubre una gran variedad de temas, desde agricultura y seguridad alimentaria hasta educación, incluyendo salud, desarrollo social, desarrollo urbano, e incluso cuestiones relacionadas con el agua. **Como temas, podrían extraerse las siguientes categorías y subcategorías** [**definidas en esta taxonomía**](https://thedocs.worldbank.org/en/doc/275841490966525495-0290022017/original/NewThemeTaxonomyanddefinitionsrevisedJuly012016.pdf)**:**

* 112 Public Expenditure policy
* 131 Inclusive growth
* 22 Jobs
* 24 Enterprise Development
* 323 MSME Finance
* 324 Financial Inclusion
* 51 Social inclusion
* 651 Access to Education
* 632 Health service delivery

Una vez recopilados los datos de noticias relevantes, podemos calcular la puntuación del tono de los artículos y la tasa de popularidad del tema promediando las medidas obtenidas de GDELT para los artículos seleccionados durante el período aplicado en la extracción de los datos.

Se propone extraer estos dos tipos de datos (Google Trends y GDELT) para posteriormente poderlos aplicar de manera simultánea en el modelo de machine learning que se sugiere a continuación. Si por cuestión de limitaciones temporales sólo uno de los tipos de datos pudiera ser extraído, los estudiantes deberán centrarse en Google Trends, dando prioridad a esta fuente de datos frente a GDELT.

### **Otros recursos**

<https://github.com/pat310/google-trends-api/wiki/Google-Trends-Categories>

<https://serpapi.com/google-trends-categories>

https://oecd.ai/en/googletrends

https://thedocs.worldbank.org/en/doc/275841490966525495-0290022017/original/NewThemeTaxonomyanddefinitionsrevisedJuly012016.pdf

**Sobre la aproximación analítica**

En comparación con los datos convencionales, Google Trends y GDELT tienen una historia corta (que comienza en 2004 y 2014, respectivamente), aunque una cobertura geográfica grande. De este modo, a pesar de la corta serie temporal, al usar datos de varios países (“panel data”) podemos generar conjuntos de datos mayores donde poder entrenar algoritmos complejos.

Para Google Trends, se extraerán datos mensuales. Estos datos se promediarán después para hacerlos anuales y que coincidan con la frecuencia de la variable a predecir. Del mismo modo, los datos diarios promediados a una frecuencia mensual de GDELT, se promediarán también a valores anuales.

Los datos de búsqueda de Google y los de GDELT tienen una dimensionalidad muy elevada, con el número de variables mucho mayor que el número de observaciones. Por lo tanto, será necesario lidiar con esta alta dimensionalidad.

Primero, se estimará un modelo anual basado en índices de búsqueda muestreados con una frecuencia mensual. En segundo lugar, el modelo se aplicará a las series mensuales de Google Trends y de GDELT para obtener predicciones mensuales de la variable respuesta. Esto implica asumir que la relación entre los datos de Google Trends y la variable “output” no dependen de la frecuencia a la que se muestreen. Este enfoque nos permite desagregar la variable de salida a una frecuencia mensual.

Las variables predictoras serán las series temporales extraídas a partir de las categorías de Google Trends, y el número de noticias por tema y tono de GDELT. Para incorporar la heterogeneidad existente en distintos países y las clases de sexo indicadas en la variable respuesta, la variable país y la variable sexo deberían también incluirse como “dummies”.

### **Recomendaciones sobre el diseño del modelo**

En este trabajo esperaríamos que muchas variables tuvieran una contribución pequeña para explicar la variable respuesta por lo que el uso de redes neuronales parece una aproximación preferible frente a otras alternativas (Random forest, XGBoost) que funcionan mejor cuando sólo unas pocas variables presentan una gran contribución de cara a explicar la variable respuesta. Además, frente a otros posibles modelos (e.g., ridge regression), las redes neuronales nos permitirán incluir relaciones no lineales entre las variables.

#### LSTM

Usaremos un **modelo de red neuronal de tipo “long–short term memory” LSTM**, que es un tipo de red neuronal recurrente (RNN).El modelo LSTM se entrenará con la secuencia de observaciones procedentes de Google Trends y/o GDELT como "inputs" con el fin de predecir la variable respuesta (*Population by sex, age, citizenship and labour status*). Además de estas variables, las variables “sexo” y “país” contenidas en la variable respuesta, se incluirán como variables “dummy”, de manera que tendremos que construir un modelo **LSTM multivariante**, lo que significa que, en este modelo, una serie de datos temporales (la variable respuesta) se modela a partir de varias series temporales de datos que se introducen como entradas (inputs) del modelo. Las series temporales de las variables input (derivadas de Google Trends y/o GDELT) son “paralelas” porque las dos series presentan observaciones en los mismos intervalos de tiempo (datos anuales). De este modo, los datos se organizarán de manera que cada fila contenga un intervalo temporal (año), y cada columna será cada una de las series temporales de las variables input.

De entre los posibles modelos LSTM (stacked, bidirectional, CNN, ConvLSTM), proponemos utilizar un **Vanilla LSTM** porque este tipo de modelo parece capturar bien las dinámicas de las series temporales y consta de una capa oculta de unidades LSTM, y una capa de salida que es la que se utiliza para hacer la predicción. Podemos definir un **modelo con 50 unidades LSTM en la capa oculta y una capa de salida que prediga un único valor numérico. El modelo puede ajustarse usando la versión eficiente de Adam del gradiente estocástico descendente y ser optimizado usando el error cuadrático medio o *MSE loss function*.** [En Python puede hacerse de la siguiente manera](https://machinelearningmastery.com/how-to-develop-lstm-models-for-time-series-forecasting/):

# define model

model = Sequential()

model.add(LSTM(50, activation='relu', input\_shape=(n\_steps, n\_features)))

model.add(Dense(1))

model.compile(optimizer='adam', loss='mse')

**Y con una función “relu”** de activación. En este modelo se incluirán las series temporales las variables de entrada (temas de Google Trends y GDELT, tono de GDELT), y las variables dummy (sexo y país) para predecir la variable respuesta. Si lo limitado de la muestra inicial (datos anuales entre 2004-2024 en el caso de Google Trends o entre 2014-2024 en el caso de GDELT) no permitiera esta elevada dimensionalidad (modelar un gran número de variables), habría que reducir el número de inputs a modelar, siguiendo las premisas indicadas en ese momento por el equipo de Randbee.

### **Otros recursos**

<https://machinelearningmastery.com/how-to-develop-lstm-models-for-time-series-forecasting/>

<https://towardsdatascience.com/exploring-the-lstm-neural-network-model-for-time-series-8b7685aa8cf>

https://www.machinelearningplus.com/time-series/time-series-analysis-python/

https://melaniesoek0120.medium.com/covid-19-global-data-time-series-prediction-with-lstm-recurrent-neural-networks-f7825c4a1f6f#:~:text=Long%20Short%2DTerm%20Memory%20(LSTM,or%20multiple%20input%20forecasting%20problems

<https://github.com/dhopp1/nowcast_lstm>

<https://github.com/dhopp1/nowcastLSTM>

https://smltar.com/dldnn

LSTM en R: **ts.lstm** package, basado en módulos de **keras y tensorflow** (<https://search.r-project.org/CRAN/refmans/TSLSTM/html/ts.lstm.html> )