



Department of Mathematics

# Navigating the Digital Landscape: An Insight into Time Series Forecasting for Search Engine Optimization

Fortunato Nucera

CID: 02165798

Supervised by Dr. Ed Cohen

September 2023

Submitted in partial fulfilment of the requirements for the  
MSc in Machine Learning and Data Science of  
Imperial College London

The work contained in this thesis is my own work unless otherwise stated.

Signed: Fortunato Nucera

Date: September 2023

# Abstract

Marketing strategies aim to boost a company's revenue by amplifying the sales volume, which is intrinsically tied to the popularity of a given product. The traditional approaches to marketing fall into two categories: active and passive. The former involves a company's direct effort to reach out to potential customers to offer a product, whereas the latter assumes that web presence and public relations efforts can successfully lead to a larger volume of sales. While none of the two categories is inherently better than the other, active marketing involves a financial burden that is either absent or limited in the passive approach. Among the passive marketing techniques, Search Engine Optimization (SEO) encompasses a wide range of methods that websites can enact to achieve a high ranking among search engine results, one of these being keyword optimization.

SEO through keyword optimization involves the identification of keywords that users are likely to search for, and the usage of those in websites and blog articles. Forecasting such keywords offers insight into future business opportunities, but time series forecasting is inherently a statistical discipline that is not normally taught in most university-level programs with a focus on marketing. In addition, even the simplest libraries for classical time series forecasting require an understanding of the statistical underpinnings for the proper hyperparameter setting.

We propose a novel methodology for time series forecasting based on Gaussian Processes (GPs) and automatic kernel selection via Widely Applicable Bayesian Information Criterion (WBIC), in addition to several hybrid methods combining GPs and Neural Networks to capture both the local and the non-local (either global or cluster-wise) behavior of the time series data. We compare these approaches with other classical methodologies requiring hyperparameter tuning and Machine Learning (ML) and Deep Learning (DL) techniques on a data set built from Google Trends which includes AI-related keywords.

While issues such as scalability and generalizability of the results to other data sets remain open, the results on the Google Trends-based data set show good predictive skills with no human intervention/interaction.

## Acknowledgements

I would like to express my profound gratitude and appreciation to all those who have played a part, both directly or indirectly, in completing this project.

First and foremost, I am immensely grateful for the invaluable guidance and unwavering support provided by my supervisor, Dr. Ed Cohen. His expertise, insightful feedback, and continuous encouragement have been instrumental in shaping this research.

I am also thankful to the faculty members and lecturers throughout the entire program for imparting their knowledge and teaching me the essential concepts that have helped me reason in a pragmatic, logical, and statistically sound manner.

Additionally, I would like to extend my heartfelt appreciation to all my classmates. Pursuing this degree remotely while simultaneously working full-time has been challenging, but their camaraderie and shared experiences have been a source of motivation and support. Knowing that I was not alone in this academic endeavor gave me the confidence and determination to overcome the obstacles along the way.

Lastly, I am thankful to my family and friends, for their unconditional love and understanding, especially during times when my workload may have caused moments of stress and unpredictability. They are my constant source of inspiration and admiration.

*Omnibus gratias ago, quorum auxilio opus hoc perfici potuit.*

## Glossary

<b>AI</b>	Artificial Intelligence.
<b>AE975</b>	The 0.975 quantile of the absolute error distribution. The value below which 97.5% of the differences between the observed and predicted values fall, without considering their direction (positive or negative).
<b>BIC</b>	Bayesian Information Criterion.
<b>DL</b>	Deep Learning.
<b>DLM</b>	Dynamic Linear Model.
<b>DTW</b>	Dynamic Time Warping.
<b>EM</b>	Expectation Maximization.
<b>GP</b>	Gaussian Process.
<b>MAPE</b>	Maximum A Posteriori Estimate or Mean Absolute Percentage Error (it depends on context).
<b>ML</b>	Machine Learning.
<b>MLE</b>	Maximum Likelihood Estimation/Estimate.
<b>MSE</b>	Mean Squared Error.
<b>RMSE</b>	Root Mean Squared Error.
<b>SEO</b>	Search Engine Optimization.
<b>SSR</b>	Sum of Squared Residuals.
<b>WBIC</b>	Widely Applicable Bayesian Information Criterion.

---

# Contents

<b>1. Introduction</b>	<b>1</b>
<b>2. Background</b>	<b>3</b>
2.1. Literature Review . . . . .	3
2.2. Basics of Time Series Analysis . . . . .	5
2.2.1. Time Series Definition . . . . .	5
2.2.2. Covariance, Correlation, and Stationarity . . . . .	7
2.2.3. Main Tasks in Time Series Analysis . . . . .	8
2.2.4. Classical Time Series Decomposition . . . . .	9
2.2.5. Time Series Operators . . . . .	9
2.2.6. Distance between Time Series . . . . .	11
<b>3. Methods and Techniques</b>	<b>13</b>
3.1. Evaluation Metrics . . . . .	13
3.1.1. In-sample Metrics . . . . .	13
3.1.2. Forecast Metrics . . . . .	15
3.2. Approaches to Time Series Forecasting . . . . .	16
3.3. Methods . . . . .	17
3.3.1. ARIMA . . . . .	18
3.3.2. Holt-Winters . . . . .	19
3.3.3. LightGBM . . . . .	19
3.3.4. LSTM . . . . .	21
3.3.5. CNN . . . . .	22
3.3.6. Gaussian Processes . . . . .	23
3.3.7. Hybrid Methods . . . . .	28
<b>4. Pre-processing</b>	<b>32</b>
4.1. Data Retrieval . . . . .	32
4.1.1. Google’s Machine Learning Glossary . . . . .	33
4.1.2. OpenAI’s ChatGPT-4 . . . . .	33
4.1.3. Web of Science . . . . .	34
4.1.4. List Merging and Data Set Creation . . . . .	34
4.2. Exploratory Data Analysis, Cleanup, and Visualization . . . . .	36
4.3. Scaling and Windowing . . . . .	36
<b>5. Results and Discussion</b>	<b>39</b>
5.1. Raw Data Set Results . . . . .	39
5.2. Smoothed Data Set Results . . . . .	41
5.3. Noise Effect . . . . .	42
5.4. Result comparison on a time series . . . . .	45
5.5. Results’ Summary . . . . .	46
5.6. Limitations . . . . .	47

<b>6. Conclusion and Further Work</b>	<b>49</b>
<b>A. Keywords Lists</b>	<b>A1</b>
A.1. Google Machine Learning Glossary . . . . .	A1
A.2. ChatGPT-4 . . . . .	A3
A.3. Web of Science . . . . .	A4

# 1. Introduction

The first few years of the third millennium have witnessed a rapid and unparalleled flourishing of computing power, algorithms, and, above everything, access to the internet. Public data from the [World Bank \(2023\)](#) shows that the global percentage of individuals connected to the internet has increased nearly 9-fold from 7% in 2000 to 63% in 2021. While not all countries allow access to the same content - owing to issues of intellectual property, desire for "cyber-sovereignty", or, more simply, censorship - the amount of information we are exposed to on a daily basis has increased exponentially, leading to what is referred to as "The Rise of the Network Society" ([Castells, 2009](#)). In relation to purchases and sales, the increase in access to the internet can be seen as a double-edged sword: on the one hand, the customer will be prompted with a larger variety of options, allowing for a more informed - and possibly better - deal; on the other, finding the right item among a myriad of similar entries can be discouraging and may reduce sales for the vendor or service provider.

It is in the context of such a dynamic and crowded internet environment that advertising becomes essential: through robust marketing strategies, the vendors may be able to obtain an edge over the competition, which can in turn trigger higher sales volume and revenue.

Although formal definitions can be ambiguous, many professionals in the industry typically divide standard advertising strategies into two broad categories: active and passive. Active advertising, consisting in pushing a message about a given product to the consumer, often operates through mailing lists or broadcasts on messaging apps, a set of practices that many customers might find highly intrusive. In contrast, passive advertising expects the consumer to notice the advertised product, thus gaining interest in it, and is usually operated via the placement of billboards, the purchase of web banners, or, in a subtler yet more impactful manner, via Search Engine Optimization (SEO).

SEO encompasses a wide range of techniques aiming to boost a page's visibility - and ranking - among search engine results. One primary factor in SEO is keyword optimization, which aims to refine the text of a web page to better match the users' searches. In this research, we focus on leveraging time series forecasting to predict the future search volumes for a specific set of keywords of interest, thus allowing marketing professionals to anticipate the public's interest in certain topics and therefore devising effective business strategies. As it is not hard to imagine some correlation between click-through rate and sales volume, keyword optimization is often considered an inexpensive way to acquire higher online visibility. However, such a perception of inexpensiveness might be misleading.

The cost of SEO is the result of two different components: a direct cost, i.e. the cost of *introducing* suitable keywords in the website, which is essentially null, and an indirect cost, or the cost of *deciding* which keywords are suitable for introduction. As the latter requires a certain amount of skilled manpower, the ensuing cost may not be negligible and can constitute an overhead in the total expense for SEO operations. Furthermore, a lengthy analysis may lead to delays, which in turn could result in missed business opportunities.

Another issue abounding in the advertising and IT industries is the shortage of skilled workforce. While text-based SEO professionals may need a solid background in statistics or machine learning to successfully employ the currently available time series forecasting models, the recent developments in AI have made hiring such professional figures an expensive process

with uncertain outcomes. Thus, the prospect of performing SEO without relying on highly skilled professionals is becoming increasingly appealing.

Furthermore, several world-renowned consulting firms appear to have a positive outlook on the future effect of AI on mankind. According to the estimates of [PwC \(2023\)](#), AI is expected to contribute up to \$15.7 trillion to the global economy by 2030, whereas [McKinsey & Company \(2023\)](#) estimates the impact of generative AI only between \$2.6 trillion and \$4.4 trillion. Although these are only estimates of the future of the AI business, they provide us with a glimpse into the potential magnitude of the influence of AI on our daily lives. This is the main reason why we have decided to focus the text-based SEO investigation on AI-related keywords.

Through this project, we wish to address the following research questions:

1. How can time series forecasting be utilized to predict the public interest in AI-related technologies?
2. How can the forecasting process be simplified to become accessible to users lacking a background in statistics and machine learning?
3. How do different models compare in terms of predictive accuracy?
4. Does scalability constitute a trade-off with model automation?

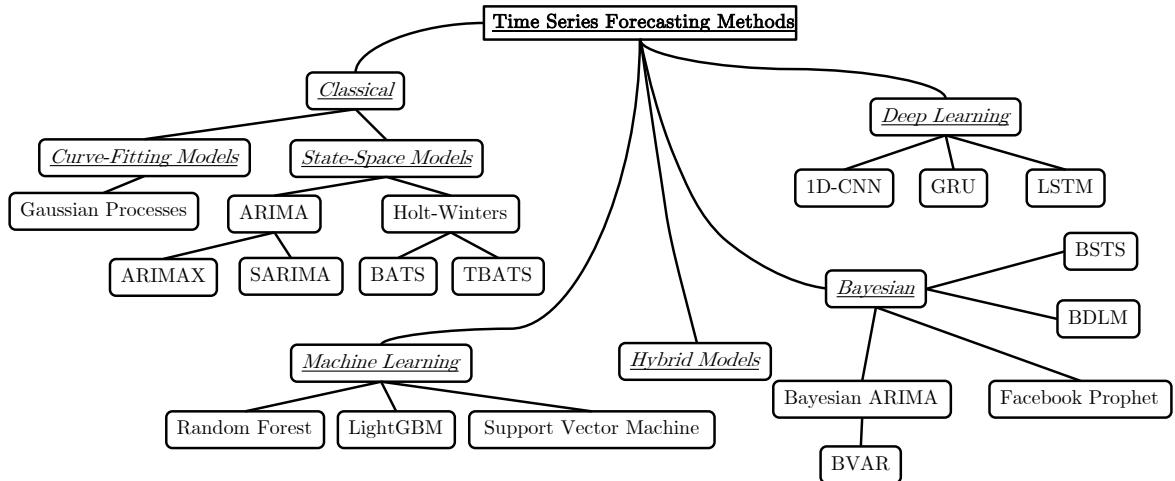
As we unravel the potential of time series forecasting for AI-related SEO, our objectives remain clear: streamlining the forecasting process, assessing the models' accuracy, discussing the scalability concerns, and empowering businesses to thrive in the ever-evolving digital marketplace.

## 2. Background

This chapter will lay the foundations for the entire project. Here we will provide a literature review and introduce the basics of time series analysis, which will be helpful for the remainder of the discussion.

### 2.1. Literature Review

Univariate time series forecasting has a long history in scientific literature, stretching across multiple disciplines, including Statistics, Machine Learning (ML), and Deep Learning (DL). This section will clarify how the models commonly used for forecasting relate to one another, and what kind of benefits and drawbacks they feature. A selection of algorithms and the family they belong to is shown in Figure 2.1.



**Figure 2.1.:** A diagram depicting common algorithms for Time Series Forecasting.

The classical approaches for time series forecasting can be subdivided into curve-fitting models and state-space models. The category of curve fitting models features, at its core, a curve fitting method. In practice, the models use a vector containing time values as covariates, and the observed time series as realization of an underlying stochastic process. The function is then fitted either via Maximum Likelihood Estimation (MLE), Expectation-Maximization (EM), Mean Squared Error (MSE) minimization, or, in other scenarios, full Bayesian inference. In the context of time series forecasting, Gaussian Processes ([Rasmussen and Williams \(2005\)](#) and [Roberts et al. \(2013\)](#)) are typically considered to be a good choice given their flexibility, robustness to outliers, and inherent easy handling of missing and unevenly sampled data. However, choosing the kernel function and dealing with their high computational cost (approximately  $O(n^3)$  where  $n$  is the number of considered time steps) are major drawbacks of their generality. Past attempts from [Duvenaud et al. \(2013\)](#), [Duvenaud \(2014\)](#), [Lloyd et al. \(2014\)](#) demonstrated the feasibility of automatic kernel selection via Bayesian Information Criterion (BIC), although such an approach bears the logical fallacy

of treating the kernel parameters as model parameters instead of *hyperparameters*. Gaussian Processes are nonparametric, thus the kernel parameters themselves do not directly affect the functional form of the data, but rather the functional form of the associated kernel. The effective number of model parameters, instead, increases with the increasing amount of available data, an issue that has not been accounted for in the cited literature.

On the other hand, state-space models include, among others, the Autoregressive Integrated Moving Average (ARIMA) model (originally in [Box et al. \(1970\)](#), but a newer description can be found in [Box et al. \(2016\)](#), [Chatfield \(2004\)](#), and [Hyndman and Athanasopoulos \(2021\)](#)) and the Holt-Winters model ([Holt \(1957\)](#) and [Winters \(1960\)](#)). The ARIMA model comprises three main building blocks: the autoregressive component, the integrated component, and the moving average component. The autoregressive component predicts future values based on past observations, while the moving average component predicts future values based on past errors. Without the integrated part, the combination of these two models would result in the ARMA model (Autoregressive Moving Average model) which assumes that the time series to be fitted is stationary. This is where the integrated part comes into play: instead of fitting ARMA on the raw time series, the model is fitted on the differenced time series, which is often stationary (refer to Section 2.2 for details on differencing). The prediction in terms of the raw time series is then retrieved through the cumulative sum. Note that, in its simplest form, ARIMA does not allow external regressors or contemplate seasonality, and cannot handle missing data. In addition, the usual assumption for ARIMA is the homoscedasticity of the residuals, which may be violated in specific contexts of variance clustering (for example, an important global event such as the onset of a pandemic may trigger anomalous search patterns, thus leading to highly volatile time series before a new state of equilibrium is reached). The ARIMA model has been extended to ARIMAX ([Hyndman and Athanasopoulos, 2021](#)), which can introduce exogenous variables to improve the predictive power of the model, and SARIMA ([Box et al., 2016](#)), which can handle seasonal patterns, in addition to VAR (Vector Autoregression) introduced in [Sims \(1980\)](#), where a set of forecasts for the time series is produced by exploiting their existing linear inter-dependencies. VAR is, however, a multivariate time series forecasting model and therefore falls outside the scope of the present project.

Another popular state-space model, the Holt-Winters model ([Holt, 1957](#))([Winters, 1960](#)), estimates trend, seasonality, and noise from the available data, and forecasts future values of the time series based on these hidden states. Holt-Winters is still extensively used in both academia and industry, thanks to its simplicity, although its uni-seasonal formulation and the inability to incorporate exogenous variables constitute its limitations. BATS (Box-Cox transformation, ARMA errors, Trend, and Seasonal components) and TBATS (Trigonometric seasonality, Box-Cox transformation, ARMA errors, Trend, and Seasonal components) ([De Livera et al., 2011](#)), extend the Holt-Winters model with multiple integer seasonalities (BATS) or non-integer seasonalities (TBATS).

The Bayesian framework features a rich literature regarding time series forecasting models. Bayesian Structural Time Series (BSTS) ([Scott and Varian, 2014](#)), much like Holt-Winters, estimates trend and seasonality under the hood, although regression covariates are introduced and selected via Bayesian model averaging. Despite its power, the main issue of BSTS is the computational cost incurred by the model averaging step. Bayesian ARIMA ([Pole et al., 2018](#)) is identical to ARIMA, but it allows the prior specification of the model parameter. The resulting model allows to leverage the user's know-how about the time series at hand, although the same specification may be detrimental as it introduces hyperparameters which, if not carefully chosen, will prevent the method from satisfactorily fitting the time series. BVAR (Bayesian Vector Autoregression) ([Baíbura et al., 2010](#)) is the Bayesian counterpart of

the VAR model, which we introduced before, and shares both strengths and weaknesses with its frequentist counterpart although, as in the case of Bayesian ARIMA, the prior elicitation may provide an edge over its competitor. Bayesian Dynamic Linear Models (DLMs) ([West and Harrison, 2006](#)), are highly flexible state-space models which can handle missing data and exogenous variables, although the requirement of state-space equation specification demands a strong statistical modeling background, thus hindering the usability in most non-academic settings. A more recent approach has been proposed in [Taylor and Letham \(2018\)](#), resulting in the release of the popular Facebook Prophet forecasting library. Prophet decomposes the time series into a sum or a product of trend, seasonality, holiday, and noise, elicits the priors on all the model's parameters, and then obtains the maximum *a posteriori* estimate using the limited-memory Broyden–Fletcher–Goldfarb–Shanno algorithm (L-BFGS), although full posterior inference is also available. Prophet is highly popular in the time series forecasting community, although the lack of options for the trend, which can only be linear or logistic, and the unsuitability for high-frequency time series, make the method less general than one would wish.

The use of ML and DL methods for time series forecasting is rather straightforward ([Bontempi et al., 2013](#)). First, a time series is split into chunks with a chosen stride, and a data set is generated so that each of the chunks is paired with a future time series value. The chunked data set can then be used to train off-the-shelf regression algorithms such as a Random Forest Regressor ([Breiman, 2001](#)), a LightGBM model ([Ke et al., 2017](#)), a Support Vector Regressor ([Cortes and Vapnik, 1995](#)), a 1-D Convolutional Neural Network ([LeCun et al., 1998](#)), a Long-Short Term Memory network ([Hochreiter and Schmidhuber, 1997](#)), or a Gated Recurrent Unit network ([Cho et al., 2014](#)). The features of some of these models, which will be employed in this project, are covered in Chapter 3. Nevertheless, [Makridakis et al. \(2018\)](#) and [Lim and Zohren \(2021\)](#) noted that ML and DL methods may suffer from issues related to inappropriate scaling and overfitting, where the latter could be mitigated through cross-training (i.e. the training of a single model for multiple time series) only if an appropriate scaling method valid across the entire data set were to be available. In addition, all the cited ML/DL models provide a prediction for the mean value of the forecast but they do not output prediction intervals, which are essential to present a clear picture of the variability in the prediction one would expect. An ancient concept - quantile regression ([Koenker and Bassett Jr, 1978](#)) - may be used to produce prediction intervals for ML/DL instead, and we will show the achieved results with such an approach in this project.

## 2.2. Basics of Time Series Analysis

This section contains the definitions and foundational ideas to be used for time series analysis. Many of the included concepts can be found, with a greater degree of detail, in [Brockwell and Davis \(2006\)](#) and [Grimmett and Stirzaker \(2001\)](#).

### 2.2.1. Time Series Definition

The topic of time series analysis borrows its statistical underpinning from the much broader field of stochastic processes. We therefore believe it is important to first define stochastic processes, and only then interpret these in the context of time series.

**Definition 2.2.1 (Stochastic Process).** A stochastic process is a family of random variables  $\{X_t : t \in T\}$  defined on a common probability space  $(\Omega, \mathcal{F}, P)$ , where  $X_t : \Omega \rightarrow \mathbb{R}$  are the random variables,  $T$  is the index set,  $\mathcal{F}$  is the  $\sigma$ -algebra and  $P$  is a probability measure.

The reader may notice that when pointing out the index set  $T$  we did not make specific assumptions on what kind of set  $T$  must be. The reason for this lies in the fact that different assumptions on the index set lead to different types of the associated stochastic processes. For example:

- if  $T$  is a time set, then the stochastic process is a *time series*.
- if  $T$  is a space set (for example the set of points on a sphere), then we refer to it as a *spatial stochastic process* (useful, for example, in modeling the turbulent fluid dynamics around a bluff body).

Within the scope of this project, we are specifically interested in time series, thus we assume  $T$  to be the time set.

**Definition 2.2.2 (Time Series).** A time series is a stochastic process  $\{X_t : t \in T\}$  where  $T$  is the time set.

For this project, we will only focus on *discrete-time* time series, and impose  $T \equiv \mathbb{Z}$ . However, *continuous-time* time series are also valid statistical entities, although it is often necessary to convert them into discrete-time time series for reasons of measurement frequency limitations or data storage/computational resources. It is also important to note that the attribute *discrete* refers to the time domain, and not to the values the time series may assume, which might be either continuous or discrete.

A more practical definition, which we will employ throughout the remainder of the project, is provided below.

**Definition 2.2.3 (Time Series (reviewed)).** A time series is a set of observations  $x \in \{x_{t_1}, x_{t_2}, \dots, x_{t_N}\}$  recorded at time  $t \in T$  where  $T$  is an ordered set  $T = \{t_1, t_2, \dots, t_N\}$ .

Importantly, the index set  $T$  used in time series is endowed with an *order*. As a result, values in a time series are not exchangeable and - as an implication - not independent. Thus, two given time series  $\mathbf{x} = \{x_t\}$  and  $\mathbf{y} = \{y_t\}$  are different unless  $x_t = y_t \quad \forall t \in T$ . Please note that we will adopt the notation using uppercase letters for the random variables and lowercase letters for their associated realizations. Additionally, the bold font will be employed to indicate a vector (of either realizations or random variables).

The underlying probability distribution of the time series is also called *finite-dimensional distribution* of the stochastic process.

**Definition 2.2.4 (Finite-Dimensional Distribution).** Let  $\{X_t : t \in T\}$  be a stochastic process indexed by the discrete ordered set  $T = \{t_1, t_2, \dots, t_N\}$ , the distribution  $X_{t_1:t_N}$  can be written as:

$$P_{t_1, t_2, \dots, t_N}^X = P(X_{t_1}, X_{t_2}, \dots, X_{t_N}) \quad (2.1)$$

Since, as previously stated, the index set  $T$  is ordered, the finite-dimensional distribution cannot, in general, be factorized into the product of the marginal distributions of the single random variables that compose it. In other words:

$$P_{t_1, t_2, \dots, t_N}^X = P(X_{t_1}, X_{t_2}, \dots, X_{t_N}) \neq \prod_{i=1}^N P(X_{t_i})$$

### 2.2.2. Covariance, Correlation, and Stationarity

In this section, the concepts of autocovariance, cross-covariance, autocorrelation, cross-correlation, and stationarity, both in its strong and weak instances, are presented.

**Definition 2.2.5 (Autocovariance).** Let  $\{X_t\}$  be a stochastic process defined on the index set  $T$ . The autocovariance  $C_{XX}(t_1, t_2)$  is the covariance <sup>1</sup> between the random variables composing the stochastic process at time steps  $t_1, t_2 \in T$ .

$$C_{XX}(t_1, t_2) = \text{cov}(X_{t_1}, X_{t_2}) \quad \text{where } t_1, t_2 \in T \quad (2.2)$$

On the other hand, the cross-covariance is simply the covariance at (possibly) distinct time steps for (possibly) distinct stochastic processes (note that the autocovariance can be seen as a special case of cross-covariance).

**Definition 2.2.6 (Cross-covariance).** Let  $\{X_t\}$  and  $\{Y_{t'}\}$  be two stochastic processes defined on possibly two distinct index sets  $T$  and  $T'$ , respectively. The cross-covariance  $C_{XY}(t_1, t'_1)$  is the covariance between the random variables composing the two stochastic processes at  $t_1 \in T$  and  $t'_1 \in T'$ .

$$C_{XY}(t_1, t'_1) = \text{cov}(X_{t_1}, Y_{t'_1}) \quad \text{where } t_1 \in T, t'_1 \in T' \quad (2.3)$$

**Definition 2.2.7 (Autocorrelation and Cross-correlation).** Let  $\{X_t\}$  and  $\{Y_{t'}\}$  be stochastic processes defined on the respective index sets  $T$  and  $T'$ . The autocorrelation is the autocovariance normalized by the standard deviation of the stochastic processes.

$$\rho_{XX}(t_1, t_2) = \frac{C_{XX}(t_1, t_2)}{\sqrt{C_{XX}(t_1, t_1)} \sqrt{C_{XX}(t_2, t_2)}} = \frac{C_{XX}(t_1, t_2)}{\sigma_{X,t_1} \sigma_{X,t_2}} \quad \text{where } t_1, t_2 \in T \quad (2.4)$$

where the standard deviation  $\sigma_{X,t} = \sqrt{C_{XX}(t, t)}$ .

Similarly, the cross-correlation is defined as:

$$\rho_{XY}(t_1, t'_1) = \frac{C_{XY}(t_1, t'_1)}{\sqrt{C_{XX}(t_1, t_1)} \sqrt{C_{YY}(t'_1, t'_1)}} = \frac{C_{XY}(t_1, t'_1)}{\sigma_{X,t_1} \sigma_{Y,t'_1}} \quad \text{where } t_1 \in T, t'_1 \in T' \quad (2.5)$$

Finally, stationarity is arguably the most descriptive property of a stochastic process (and, by extension, of a time series). This property can be either *strong* or *weak*. Notably, many models used for time series forecasting require the stationarity of the underlying time series.

**Definition 2.2.8 (Strong Stationarity).** Let  $\{X_t : t \in T\}$  be a stochastic process indexed by  $T$  and  $h$  be a shift such that if a subset  $\{t_1, t_2, \dots, t_n\} \subseteq T$  then  $\{t_{h+1}, t_{h+2}, \dots, t_{h+n}\} \subseteq T$ . The stochastic process is strictly stationary if its finite-dimensional distribution  $P_{t_1, t_2, \dots, t_n}^X$  is constant under the shift  $h$ . In other words:

$$P_{t_1, t_2, \dots, t_n}^X = P_{t_1+h, t_2+h, \dots, t_n+h}^X \quad (2.6)$$

This requirement is often too restrictive for the stochastic process to be of any practical use. For this reason, there exists another formulation of stationarity, namely *weak* stationarity, that relaxes the probabilistic requirements.

---

<sup>1</sup>Recall that the covariance of two random variables  $X$  and  $Y$  is  $\text{cov}(X, Y) = E[(X - E[X])(Y - E[Y])]$  or  $\text{cov}(X, Y) = E[XY] - E[X]E[Y]$  with  $E[\cdot]$  indicating the expected value of the random variable.

**Definition 2.2.9 (Weak Stationarity).** Let  $\{X_t : t \in T\}$  be a stochastic process indexed by  $T$ ,  $t_1, t_2 \in T$ , and  $h \in \mathbb{Z}$  be a lag such that  $t + h \in T$ . A stochastic process is weakly stationary if:

$$\mathbb{E}[X_{t_1}] = \mathbb{E}[X_{t_2}] = c < \infty \quad (2.7)$$

$$C_{XX}(t_1, t_2) = C_{XX}(t_1 + h, t_2 + h) = C_{XX}(0, h) < \infty \quad (2.8)$$

where  $\mathbb{E}[\cdot]$  represents the expectation operator,  $c$  is a constant, and  $C_{XX}(t_1, t_2)$  is the autocovariance function. In other words, a weakly stationary stochastic process features a finite and constant first-order moment and a finite covariance that only depends on the lag  $h$ : neither the first-order nor the second-order moments are allowed to depend on the time index. From now on, we will use the term *stationarity* to indicate *weak stationarity* unless stated otherwise.

### 2.2.3. Main Tasks in Time Series Analysis

So far, we have discussed concepts that are largely borrowed from the theory of stochastic processes. The next section will focus on applying these concepts to analyze and forecast time series data.

Historically, scientists involved in time series analysis have dedicated significant resources to the tasks of forecasting, correlation structure estimation, and frequency domain analysis. However, with the increased availability of computational power at modest prices, thanks also to affordable cloud environments, the focus of interest - especially among the community bridging statisticians and machine learning scientists - has gradually shifted.

In particular, much of the attention is now paid to forecasting, clustering, classification, and anomaly detection (although the last task can be seen either as separate or as a specific application of the classification/clustering tasks). The objectives for each task are listed below:

- **forecasting:** the forecasting task aims to predict the future values of a time series given its historical observations. In mathematical terms, we want to predict the following conditional distribution:

$$P(X_{t_N+1}, X_{t_N+2}, \dots, X_{t_N+M} | X_{t_1}, X_{t_2}, \dots, X_{t_N})$$

where  $M$  is the *forecast horizon*. A typical use case of forecasting may be the prediction of the temperature on a given day or the prediction of the normalized search volume for a keyword on a search engine (the topic of this research).

- **classification:** the classification task aims to predict, given a set of  $B$  distinct time series each labeled as one of  $K$  classes, the class of an unseen example whose label is unknown. The data set may be written as:

$$\left\{ \left( \mathbf{x}^{(1)}, c^{(1)} \right), \left( \mathbf{x}^{(2)}, c^{(2)} \right), \dots, \left( \mathbf{x}^{(B)}, c^{(B)} \right) \right\} \text{ where } c^{(i)} \in \{1, \dots, K\}$$

Time series classification is especially useful in the context of Electroencephalography (EEG) interpretation (Coelho et al., 2017) and cardiac arrhythmias detection (Jovic and Jovic, 2017).

- **clustering:** the purpose of the clustering task is to detect, given unlabeled time series, the latent groupings those time series belong to. The power of clustering lies in the

fact that the data set does not need to be labeled in order to draw insights from it. In addition, in specific circumstances, if the underlying data manifold is properly learned, new examples may be generated from the latent groupings (two examples of generative models are the Variational Auto Encoder from Kingma and Welling (2013) and the Gaussian Process Latent Variable Model from Lawrence (2003)). Most of the traditional clustering algorithms such as K-Means, K-Medoids, DBSCAN, OPTICS, and DENGLUE can be used for time series clustering (a detailed compendium for these can be found in Han et al. (2011)). However, all of the clustering methods rely on the definition of a distance metric, which is often chosen to be Dynamic Time Warping (DTW). More information on such metric is provided in Section 2.2.13.

#### 2.2.4. Classical Time Series Decomposition

The time series decomposition framework is general but somewhat inflexible. A detailed description of it can be found in Hyndman and Athanasopoulos (2021).

Without loss of generality, a time series  $\{X_t\}$  can be decomposed into three distinct elementary time series archetypes:

- **trend**  $m_t$ : a slowly increasing/decreasing time series (which is sometimes assumed to be linear in order to simplify the model). The adverb "slowly" is to be interpreted in reference to the global time scale of the time series.
- **seasonality**  $s_t$ : a component exhibiting fluctuating patterns with constant or variable magnitude.
- **noise**  $Y_t$ : a random variable incorporating the residual that is left unaccounted for by both trend and seasonality components.

These archetypes can be combined together using either addition (Equation 2.9) or multiplication (Equation 2.10).

$$X_t = m_t + s_t + Y_t \quad (2.9)$$

$$X_t = m_t \times s_t \times Y_t \quad (2.10)$$

In particular, the additive form is appropriate when the magnitude of the oscillation about the trend is approximately constant across the considered time span. In other circumstances, multiplicative decomposition is recommended instead.

It is evident from the discussion above that, according to the definitions, a time series featuring either a trend or a seasonal component cannot be considered stationary per Definition 2.2.9. In fact, if trend  $m_t$  and seasonality  $s_t$  were to be present then  $X_t = m_t + s_t + Y_t$  and even if the noise  $Y_t \sim \mathcal{N}(0, \sigma^2)$ , then  $E[X_t] = E[m_t + s_t]$ . Since both  $m_t$  and  $s_t$ , in general, may depend on the considered index  $t$ , then  $\exists t_1, t_2$  such that  $E[X_{t_1}] \neq E[X_{t_2}]$ , thus failing to satisfy the finite and constant first-order moment condition for weak stationarity.

#### 2.2.5. Time Series Operators

The appropriate use of time series operators allows us to write the equations for the forecasting models more succinctly.

**Definition 2.2.10 (Backshift Operator).** The backshift operator  $B$  maps the value of a time series at time  $t$  to the value of the same time series at time  $t - 1$ .

$$Bx_t := x_{t-1} \quad (2.11)$$

The backshift operator  $B$  also allows for a compact reference to several time steps in the past. For example:

$$B^n x_t = B(B^{n-1} x_t) = B(B(B^{n-2} x_t)) = B(\dots B x_t) = B x_{t-n+1} = x_{t-n}$$

$B$  will be particularly useful in the description of the Autoregressive Model (AR).

The second operator we introduce is the differencing operator  $\Delta$ . Please note that both the nabla symbol  $\nabla$  and the Laplacian symbol, or capital Greek letter  $\Delta$ , are used in literature. Here we adopt the latter notation.

**Definition 2.2.11** (*Difference Operator*). The difference operator  $\Delta$  maps the value of a time series at a time step  $t$  to the difference between the time series value at time steps  $t$  and  $t - 1$ .

$$\Delta x_t := x_t - x_{t-1} = x_t - B x_t = (1 - B)x_t \quad (2.12)$$

Note that the operator can be generalized to a seasonal differencing of period  $s$  as:

$$\Delta_s x_t := x_t - x_{t-s} = x_t - B^s x_t = (1 - B^s)x_t \quad (2.13)$$

The generalized form of the difference operator is especially useful when a seasonal component of period  $s$  is present, as an  $s$ -period differencing would help mitigate or remove the seasonal pattern without resorting to high-order models.

Differencing a time series before fitting a machine learning/statistical model may lead to some benefits as well as some drawbacks. While stationarity is the requirement of the Autoregressive Moving Average (ARMA) model, differencing involves the loss of information of one time step for each differencing iteration, which may be particularly problematic when the available time series to forecast is rather short. In addition, differencing may affect the interpretability of the forecast, which will need to be integrated in order to be actionable, and over-differencing may introduce unnecessary complexity to the model, leading to possible overfitting. Additionally, assessing the appropriate differencing order can be achieved through the application of the Augmented Dickey-Fuller test (Dickey and Fuller, 1979). The mathematical derivation of this hypothesis test is quite involved and requires the analysis of the  $t$ -statistic of the coefficient of the first-order difference in an autoregressive model fitted on the time series. The test can be carried out using freely available libraries such as `statsmodels` in Python and we, therefore, provide only the test's hypotheses, without some details which would otherwise make the discussion verbose.

**Definition 2.2.12** (*Augmented Dickey-Fuller*). Let  $\{X_t\}$  be a time series defined on an index set  $T$ .

- **Null Hypothesis:**  $H_0$ ,  $\{X_t\}$  is non-stationary.
- **Alternative Hypothesis:**  $H_1$ ,  $\{X_t\}$  is stationary

at a confidence level  $\alpha$ . If the  $p$ -value of the test falls below the confidence level  $\alpha$  threshold, then the time series is deemed to be stationary.

Importantly, the Augmented Dickey-Fuller test exhibits low power, thus resulting in many false negatives (difficulty in detecting stationarity even when the time series is actually stationary).

### 2.2.6. Distance between Time Series

The comparison of time series requires the definition of a suitable distance metric. Given two time series  $\mathbf{x} = [x_1, x_2, \dots, x_N]$  and  $\mathbf{y} = [y_1, y_2, \dots, y_N]$ , the Euclidean distance may be used:

$$d_{\text{euclidean}}(\mathbf{x}, \mathbf{y}) = \sqrt{\sum_{t=1}^{t_n} (x_i - y_i)^2}$$

This method is effective when both original time series are sampled at matching time steps. However, in general, if two time series were to be exactly the same except for a given lag, they would still be regarded as different based on the Euclidean distance. In addition, the Euclidean distance can only be defined for two time series of the same length, which may limit its practical use. Thus, a more general approach is required.

**Definition 2.2.13 (Dynamic Time Warping (DTW)).** Dynamic Time Warping ([Sakoe and Chiba, 1978](#)) was initially introduced for tasks concerning speech recognition, but its simplicity has made it popular in the field of signal processing at large.

Given two time series, DTW finds the best index pairing fulfilling the following conditions:

- Each index from one of the time series must be paired with at least one index from the other.
- The indices at the extremities must match.
- The indices cannot self-fold (i.e. the indices within a series cannot be paired with preceding indices of the other).

The algorithm for the calculation of the DTW can be easily implemented.

Let  $\mathbf{x} = [x_1, x_2, \dots, x_T]$  and  $\mathbf{y} = [y_1, y_2, \dots, y_{T'}]$  be two observed time series, and let  $D$  be an initialized matrix  $D \in \mathbb{R}^{(T+1) \times (T'+1)}$  such that:

$$D = \begin{bmatrix} 0 & \infty & \infty & \cdots \\ \infty & 0 & 0 & \cdots \\ \infty & 0 & 0 & \cdots \\ \vdots & \vdots & \vdots & \ddots \end{bmatrix}$$

Then the matrix  $D$  can be completed recursively:

$$D(t, t') = |x_t - y_{t'}| + \min\{D(t-1, t'), D(t, t'-1), D(t-1, t'-1)\} \quad (2.14)$$

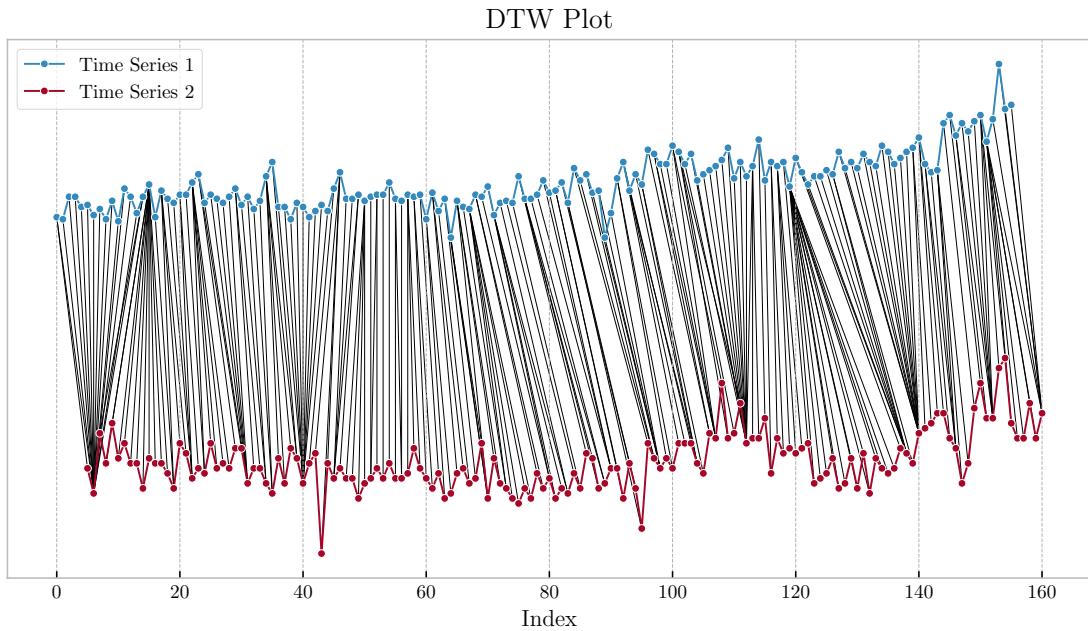
which means we always choose the path of minimum cost between  $\mathbf{x}$  and  $\mathbf{y}$ . In particular:

$$d_{\text{DTW}}(\mathbf{x}, \mathbf{y}) = D(T, T')$$

represents the DTW distance. A visualization of the index pairing for two time series can be found in [Figure 2.2](#).

What we have presented is only *one* of the possible DTW implementations, and arguably the simplest one. In fact, while some textbooks and papers employ the squared difference instead of the absolute value in [Equation 2.14](#), others put some constraints on the maximum distance between two matching indices in the two series. These modifications do not necessarily bring benefits in terms of accuracy, add an additional layer of complexity to the analysis, and are therefore excluded from the current investigation.

Finally, please note that, while it is common to refer to DTW as a distance metric, this does not satisfy *all* the requirements of such a mathematical entity: even though non-negativity and symmetry are always satisfied, the triangle inequality is not, thus the transitive property does not apply. Nevertheless, DTW is a valuable tool commonly utilized in the industry which we will apply extensively later on in the context of hybrid modeling for time series forecasting (Section 3.3.7).



**Figure 2.2.:** A representation of DTW for two standard time series.

## 3. Methods and Techniques

In this chapter, we will cover the metrics to be used for model evaluation, as well as the models to be fitted on the data set.

### 3.1. Evaluation Metrics

The proper choice of the model evaluation metrics is essential for a fair assessment of its capabilities. We commonly employ different metrics for the in-sample data and the forecast.

#### 3.1.1. In-sample Metrics

The in-sample metrics need to be able to limit the model overfitting by penalizing, either directly or indirectly, the number of model parameters. For this reason, we will resort to two different approaches depending on the ease with which the likelihood can be calculated.

**Definition 3.1.1** (*Bayesian Information Criterion*). The Bayesian Information Criterion (BIC), originally introduced in [Schwarz \(1978\)](#), defines the following quantity:

$$\text{BIC} = k \log n - 2\ell(\hat{\theta}_{\text{MLE}}) \quad (3.1)$$

where  $k$  is the number of model parameters,  $n$  is the number of data points,  $\ell$  is the log-likelihood of the data under the chosen model,  $\theta$  are the model parameters and,  $\hat{\theta}_{\text{MLE}}$  are the parameters' estimates at the maximum likelihood. The best model, based on this metric, is the one minimizing Equation 3.1.

For a fixed amount of data, a small BIC implies a large log-likelihood  $\ell$  and a small number of parameters  $k$ . As the number of parameters increases, the BIC penalizes the model for a quantity proportional to the logarithm of the number of data points. The key assumption is therefore that the number of parameters of the model is well defined, which in turn implies that the underlying model must be parametric. If the model is nonparametric, the number of parameters is a function of the data, which means that the number of parameters  $k$  may not be properly defined. In order to overcome the parametric limitation, we can resort to the much newer concept of *Widely Applicable Bayesian Information Criterion*.

**Definition 3.1.2** (*Widely Applicable Bayesian Information Criterion*). The Widely Applicable Bayesian Information Criterion (WBIC), introduced in [Watanabe \(2013\)](#), defines the following quantities:

$$\begin{aligned} L_n(\theta) &= -\frac{1}{n} \sum_{i=1}^n \log p(X_i|\theta) = -\frac{1}{n} \ell(\theta) && \text{average negative log-likelihood} \\ \beta &= \frac{1}{\log(n)} && \text{inverse temperature} \\ \text{WBIC} &= \mathbb{E}_\theta^\beta [nL_n(\theta)] = \frac{\int nL_n(\theta) \prod_{i=1}^n p(X_i|\theta)^\beta \phi(\theta) d\theta}{\int \prod_{i=1}^n p(X_i|\theta)^\beta \phi(\theta) d\theta} && \text{metric} \end{aligned}$$

where  $n$  is the number of data points,  $\theta$  is the vector of trainable parameters,  $p(X_i|\theta)$  is the model likelihood,  $\phi(\theta)$  is the prior distribution of the model parameters.

One way to interpret WBIC is as the expectation of the negative log-likelihood with respect to the posterior distribution of the parameters. In other words, defining  $\pi(\theta) = \phi(\theta|X)$  as the parameters' posterior distribution obtained via Bayes' theorem, WBIC becomes:

$$\text{WBIC} = \mathbb{E}_\theta^\beta [nL_n(\theta)] = \frac{\int nL_n(\theta) \prod_{i=1}^n p(X_i|\theta)^\beta \phi(\theta) d\theta}{\int \prod_{i=1}^n p(X_i|\theta)^\beta \phi(\theta) d\theta} = \int nL_n(\theta) \pi(\theta) d\theta$$

The only problem is that the likelihood and the prior over the parameters are not necessarily conjugate, therefore the integral  $\int \prod_{i=1}^n p(X_i|\theta)^\beta \phi(\theta) d\theta$  defining the model evidence is not tractable, which is the main reason why the WBIC is not a very popular metric. In order to overcome the tractability issue in Bayesian statistics, two distinct approaches may be employed:

- resorting to Hamiltonian Markov Chain Monte Carlo (HMC) (Neal et al., 2011), involving the analysis of the samples' history in order to determine whether the chains have converged, where the convergence, in general, becomes slower as the number of model's parameters increases.
- using the Laplace approximation (Laplace, 1814) to the posterior distribution, carrying two distinct problems in itself: the slow calculation of the Hessian at the *Maximum A Posterior Estimate* and the fact that if such estimate is close to the boundaries of the parameter space, then the assumption of posterior normality (which the Laplace approximation is based on) may be unsuitable.

In this research, we propose, instead, a Monte Carlo estimate for the WBIC. Starting from the likelihood, we can write:

$$\prod_{i=1}^n p(X_i|\theta)^\beta = \exp \left\{ \log \prod_{i=1}^n p(X_i|\theta)^\beta \right\} = \exp \left\{ \beta \sum_{i=1}^n \log p(X_i|\theta) \right\} = e^{\beta \ell(\theta)}$$

Next, we can rewrite  $nL_n(\theta)$ :

$$nL_n(\theta) = - \sum_{i=1}^n \log p(X_i|\theta) = -\ell(\theta)$$

Therefore the Monte Carlo approximation for the WBIC,  $\hat{\text{WBIC}}$ , can be rewritten as:

$$\hat{\text{WBIC}} = \frac{\mathbb{E}_{\theta \sim \phi(\theta)}^\beta [-\ell(\theta_j) e^{\beta \ell(\theta_j)}]}{\mathbb{E}_{\theta \sim \phi(\theta)}^\beta [e^{\beta \ell(\theta_j)}]} = \left[ \frac{\sum_{j=1}^M -\ell(\theta_j) e^{\beta \ell(\theta_j)}}{\sum_{j=1}^M e^{\beta \ell(\theta_j)}} \right]$$

Where the numerator and the denominator of the estimate are calculated on the same samples from the prior  $\phi(\theta)$ . This approximation is also particularly useful as it allows both the numerator and the denominator to be computed through the log-sum-exp trick, thus

avoiding possible numerical stability issues. In fact:

$$\hat{WBIC} = \left[ \frac{\sum_{j=1}^M -\ell(\theta_j) e^{\beta\ell(\theta_j)}}{\sum_{j=1}^M e^{\beta\ell(\theta_j)}} \right] = \quad (3.2)$$

$$= \exp \{ \text{logSumExp}_j [\beta\ell(\theta_j) + \log(-\ell(\theta_j))] - \text{logSumExp}_j [\beta\ell(\theta_j)] \} \quad (3.3)$$

stabilizes the WBIC computation. For the purpose of this research, we will employ 100 Monte Carlo samples (thus  $M = 100$ ).

Additionally, for the Machine Learning and Deep Learning models, we will resort to the quantile (or pinball) loss, introduced originally in [Koenker and Bassett Jr \(1978\)](#) but simplified in [Chung et al. \(2021\)](#), which allows us to fit a model and output prediction intervals over the forecast.

**Definition 3.1.3 (Pinball Loss).** Let  $y_i$  be the ground truth for a data set  $\{(\mathbf{X}_i, y_i)\}_{i=1}^n$ , and  $\hat{y}_i$  be the prediction of a model  $\mathcal{M}$  such that  $\hat{y}_i = \mathcal{M}(\mathbf{X}_i)$ , then the pinball (or quantile) loss for the quantile  $\tau > 0$  is defined as:

$$\rho_\tau = \frac{1}{n} \sum_{i=1}^n (\hat{y}_i - y_i) (\mathbb{I}\{y_i \leq \hat{y}_i\} - \tau)$$

The interpretation of this loss is straightforward. For example:

- if  $\tau = 0.5$ , then  $\rho_{0.5}(y_i, \hat{y}_i) = 0.5|y_i - \hat{y}_i|$ , which is essentially the mean absolute error.
- if  $\tau = 0.95$ , then  $\rho_{0.95}(y_i, \hat{y}_i) = 0.05(\hat{y}_i - y_i)$  if  $\hat{y}_i \geq y_i$  and  $0.95(y_i - \hat{y}_i)$  otherwise. This means that the loss penalizes heavily a prediction below the ground truth.
- if  $\tau = 0.05$ , then  $\rho_{0.05}(y_i, \hat{y}_i) = 0.95(\hat{y}_i - y_i)$  if  $\hat{y}_i \geq y_i$  and  $0.05(y_i - \hat{y}_i)$  otherwise. This means that the loss penalizes heavily a prediction above the ground truth.

The pinball loss is very useful and will be extensively utilized with ML/DL models.

### 3.1.2. Forecast Metrics

As our forecast metrics, we utilize the Mean Absolute Error (MAE), the Mean Absolute Percentage Error (MAPE), and the 97.5% quantile of the Absolute Error (AE975). Specifically, the MAE gives equal weight to all errors, irrespective of their magnitude; the MAPE weighs errors in proportion to the value of the ground truth; the AE975 offers an estimate of the upper bound of the absolute errors produced by the model.

**Definition 3.1.4 (Mean Absolute Error).** Let  $y_i$  be the ground truth for a data set  $\{(\mathbf{X}_i, y_i)\}_{i=1}^n$ , and  $\hat{y}_i$  be the prediction of a model  $\mathcal{M}$  such that  $\hat{y}_i = \mathcal{M}(\mathbf{X}_i)$ , then the Mean Absolute Error (MAE) is defined as:

$$\text{MAE} = \frac{1}{n} \sum_{i=1}^n |\hat{y}_i - y_i|$$

**Definition 3.1.5 (Mean Absolute Percentage Error).** Let  $y_i$  be the ground truth for a data set  $\{(\mathbf{X}_i, y_i)\}_{i=1}^n$ , and  $\hat{y}_i$  be the prediction of a model  $\mathcal{M}$  such that  $\hat{y}_i = \mathcal{M}(\mathbf{X}_i)$ , then the Mean Absolute Percentage Error (MAPE) is defined as:

$$\text{MAPE} = \frac{1}{n} \sum_{i=1}^n \frac{|\hat{y}_i - y_i|}{|y_i|}$$

**Definition 3.1.6 (Absolute Error 97.5% Quantile).** Let  $y_i$  be the ground truth for a data set  $\{(\mathbf{X}_i, y_i)\}_{i=1}^n$ , and  $\hat{y}_i$  be the prediction of a model  $\mathcal{M}$  such that  $\hat{y}_i = \mathcal{M}(\mathbf{X}_i)$ . Define the absolute errors  $e_i = |\hat{y}_i - y_i|$ . Let  $F_E(e)$  represent the cumulative distribution function (CDF) of the absolute errors across the dataset. Then, the Absolute Error 97.5% Quantile (AE975) is defined as:

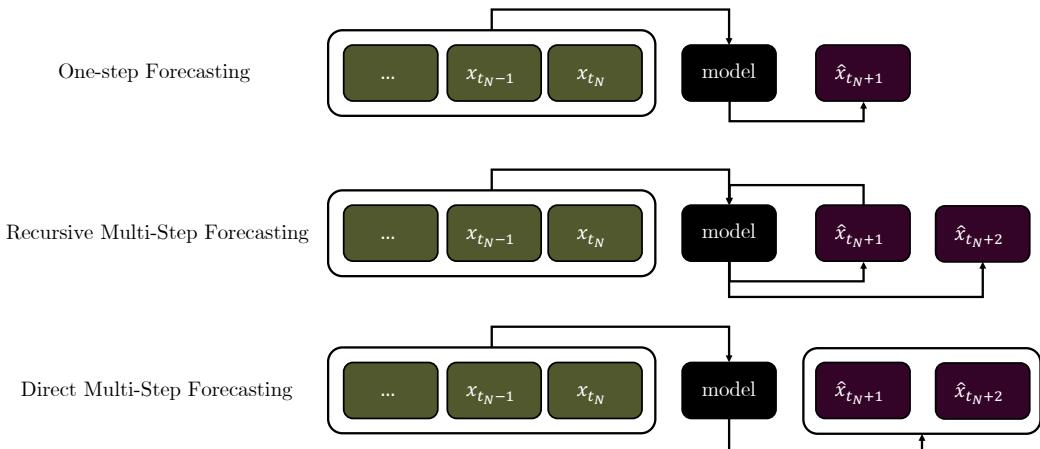
$$\text{AE975} = F_E^{-1}(0.975)$$

where  $F_E^{-1}$  is the inverse of the CDF, often referred to as the quantile function.

Note that the AE975 effectively trims the upper tail of the error distribution, thus providing a value that is less affected by outliers.

### 3.2. Approaches to Time Series Forecasting

Training a model for time series forecasting requires, in some circumstances, the prior specification of the *forecast horizon*, or the number of time steps ahead we wish to forecast the time series values for. Several forecasting strategies exist, and they are summarized in Figure 3.1.



**Figure 3.1.:** A scheme of popular time series forecasting approaches. For ease of representation, the maximum horizon is set to 2.

**Definition 3.2.1 (One-step Forecasting).** Let  $\{X_t\}$  be a time series, where  $t \in \{t_1, \dots, t_N\}$ , one-step forecasting involves the calculation of an estimate for  $X_{t+1}$  conditionally on all the available data:

$$\hat{X}_{t+1} \quad \text{given} \quad X_{t_1}, \dots, X_{t_N}$$

In most cases of practical relevance, we may wish to provide estimates for some time steps further in the future and this is where the multi-step forecasting procedure comes into play.

**Definition 3.2.2 (Recursive Multi-Step Forecasting).** Let  $\{X_t\}$  be a time series, where  $t \in \{t_1, \dots, t_N\}$ , recursive multi-step forecasting for a horizon  $K$  involves the calculation of the estimates for  $X_{t+1}$  conditionally on all the available data, and then recursively computing values further in the future by plugging in the estimates and assuming them to

be ground truth for the further forecast.

$$\begin{aligned}\hat{X}_{t_N+1} &\text{ given } X_{t_1}, \dots, X_{t_N} \\ \hat{X}_{t_N+2} &\text{ given } X_{t_1}, \dots, X_{t_N}, \hat{X}_{t_N+1} \\ &\dots \\ \hat{X}_{t_N+K} &\text{ given } X_{t_1}, \dots, X_{t_N}, \hat{X}_{t_N+1}, \dots, \hat{X}_{t_N+K-1}\end{aligned}$$

**Definition 3.2.3** (*Direct Multi-Step Forecasting*). Let  $\{X_t\}$  be a time series, where  $t \in \{t_1, \dots, t_N\}$ , direct multi-step forecasting for a horizon  $K$  involves fitting one or more models which *simultaneously* output  $K$  distinct outputs, one for each time step in the horizon:

$$\hat{X}_{t_N+1}, \dots, \hat{X}_{t_N+K} \text{ given } X_{t_1}, \dots, X_{t_N}$$

Recursive and direct multi-step forecasting are fundamentally different: the former, by plugging in estimates of the future values of the time series, leads to the propagation of errors. For some models, such as ARIMA and Holt-Winters, this is the most natural way of handling a multi-step horizon. The recursive method can also be applied, without specific countermeasures, to any Machine Learning or Deep Learning model. On the opposite, the direct multi-step forecasting allows the simultaneous prediction of several steps in the future, as long as the model allows multiple outputs (in such cases, the model parameters might be shared across the multi-step predicting head). However, some other models may only output a single value for each - possibly multidimensional - input. In such a scenario,  $K$  models can be simultaneously trained to predict  $\hat{X}_{T+1}$  to  $\hat{X}_{T+K}$  steps, although please note that, in this case, the model's parameters are not shared and the extra model complexity may cause overfitting.

### 3.3. Methods

In this section, we describe the time series forecasting methods used for comparison. In particular, we cover:

- Classical Methods: ARIMA and Holt-Winters
- Tree-Based Methods: LightGBM
- Deep Learning Methods: 1D-CNN and LSTM
- Gaussian Processes
- Hybrid Methods (Gaussian Process + Deep Learning)

A comprehensive discussion of the data set in use can be found in Chapter 4. For the current context, it is enough to mention that our selected data set comprises 281 weekly time series, each spanning 60 weeks, with 44 weeks being utilized for training and the remaining 16 weeks constituting the test set. This aspect must hint at the inherent difficulty in the inclusion of a seasonal component, which may be problematic for the following three reasons:

- a yearly seasonality, should it exist, would not be detected as the model is only exposed to 44 weeks of data (and therefore less than a year).
- automatic seasonality detection algorithms will have high uncertainty around the seasonality period.

- seasonality may include extra model complexity, which will, in turn, make the model more prone to overfitting the training data noise.

The automatic ARIMA model selection routine incorporates seasonality, whereas Holt-Winters excludes it due to the absence of a defined likelihood, precluding the application of an information criterion.

### 3.3.1. ARIMA

The ARIMA (AutoRegressive Integrated Moving Average) model ([Chatfield, 2004](#)) ([Hyndman and Athanasopoulos, 2021](#)) includes:

- an autoregressive (AR) component, which is a linear combination of prior values of the time series. The order of the autoregressive component, commonly indicated with  $p$ , determines the model's "memory" or the oldest value of the time series accounted for in the regression.
- an integrated (I) component of order  $d$ , indicating how many times the time series has been differenced prior to the fitting step.
- a moving average (MA) component of order  $q$ , expressing the linear combination of past errors to forecast future values.

Let  $\{X_t\}$  be a time series, and let  $\{Y_t\}$  be its associated  $d$ -differenced time series (see definition [2.2.11](#)). That is:

$$Y_t = \Delta^d X_t = (1 - B)^d X_t$$

then the ARIMA( $p, d, q$ ) model of time series  $\{X_t\}$  can be written as the ARMA( $p, q$ ) of  $\{Y_t\}$ :

$$Y_t = a + \sum_{i=1}^p \phi_i Y_{t-i} + \sum_{j=1}^q \theta_j \epsilon_{t-j} + \epsilon_t$$

where  $a$  is a constant (depending on the author and on the type of time series at hand, this parameter may or may not be included in the model),  $\phi_i$  are the parameters associated with the autoregressive component,  $\theta_j$  are the parameters related to the moving average component, and  $\{\epsilon_t\}$  is a zero-mean white noise process.

**Definition 3.3.1** (*White Noise Process*.). Let  $\{X_t\}$  be a time series, then  $\{X_t\}$  is a white noise process if:

- $E[X_t] = \mu$ , constant mean.
- $C_{XX}(t, t) = \sigma$ , constant variance.
- $C_{XX}(t, s) = 0$  if  $t \neq s$ , no correlation at lag larger than 0.

We will not go into the details of the model's invertibility and stationarity, as comprehensive manuals exist on the topic. It suffices to specify that, in the context of the current study, we operate the following automatic model selection procedure:

- differencing order  $d$ : the differencing order is set to be the minimum number of times the time series needs to be differenced in order to reject the null hypothesis of the Augmented Dickey-Fuller test ([Dickey and Fuller, 1979](#)), described in the Hypothesis Test [2.2.12](#).

- autoregressive order  $p$  and moving average order  $q$ : they are selected so that the BIC of the resulting model is minimized. Please note that the BIC calculation requires knowledge of the log-likelihood of the residuals, which in turn requires an assumption of the statistical model that such residuals should follow. In this study, we assume that the residuals are normally distributed, therefore the log-likelihood for the ARIMA model can be written as:

$$\ell = -\frac{n}{2} \log(2\pi) - n \log(\sigma) - \frac{1}{2\sigma^2} \sum_{t=1}^T \epsilon_t^2$$

In addition, note that if the intercept parameter  $a$  is included in the count of the total number of the model's parameters, the BIC can be rewritten as:

$$\text{BIC}_{\text{ARIMA}} = (p + q + 1) \log(T) - 2\ell \quad \text{with } a$$

or:

$$\text{BIC}_{\text{ARIMA}} = (p + q) \log(T) - 2\ell \quad \text{without } a$$

As the search space is potentially infinite, we set constraints on the orders of  $p$  and  $q$  to a total of 44, thus implying that the maximum number of model's parameters is constrained to be at most as large as the number of time steps in the series.

### 3.3.2. Holt-Winters

The Holt-Winters model ([Holt \(1957\)](#) and [Winters \(1960\)](#)) can be applied in its additive or multiplicative formulation. However, the multiplicative formulation applies the additional requirement on the positivity of the time series. Unfortunately, this requirement cannot be satisfied given that the scaling we will discuss in Section 4.3 maps the raw data to the entire real line. As a result, for the current study, we will resort only to the additive formulation of the method. Furthermore, for the reasons explained in Section 3.3, the seasonality component will be disregarded.

Let  $\{X_t\}$  be a time series, then the additive and non-seasonal Holt-Winters model, also referred to as Holt's Linear Model, can be expressed via three distinct equations:

$$\begin{aligned} \text{forecast equation: } & X_{t+h} = l_t + hb_t \\ \text{level equation: } & l_t = \alpha X_t + (1 - \alpha)(l_{t-1} + b_{t-1}) \\ \text{trend equation: } & b_t = \beta(l_t - l_{t-1}) + (1 - \beta)b_{t-1} \end{aligned}$$

where  $\alpha$  and  $\beta$  are the model's parameters to be set by minimizing the sum of squared residuals (SSR),  $h$  is the index of the forecast horizon, and the initial values for the level  $l$  and the trend  $b$  are set so that  $l_1 = X_1$  and  $b_1 = X_2 - X_1$ . This model is particularly appreciated for its simplicity and inherent estimation of the trend. Moreover,  $\alpha$  and  $\beta$  can be seen as the smoothing parameters for the level and the trend, respectively (note how these are, in fact, the coefficients of a convex combination).

### 3.3.3. LightGBM

LightGBM ([Ke et al., 2017](#)) is a recent tree-based method used for both classification and regression problems. A foundational introduction to the boosting principles can be found in [Witten et al. \(2021\)](#). In this work, we provide the general idea of tree-based algorithms,

drawing comparisons with Random Forest - a method we decided not to employ owing to both its computational complexity and time inefficiencies relative to LightGBM - the logic behind boosting, and how LightGBM speeds up the fitting step in comparison with the vanilla tree-boosting implementation.

The general idea behind regression tree algorithms is to partition the feature space via binary splits and to assign the mean value of the targets to the data points falling in the same partition. There are several criteria to determine the tree depth and how to operate the split in the feature space, although for regression they typically rely on the reduction of variance for the target variable. Although easily interpretable, regression trees are high-variance models, which tend to overfit the training set and are therefore not commonly used on their own.

The Random Forest algorithm (Breiman, 2001) is based on the idea that models' ensembles are capable to reduce the model's overall variance, thus allowing better out-of-sample performance. In particular, the Random Forest algorithm builds several trees, each of which is built on a bootstrapped version of the training set and where each of the splits is constrained to be operated on a random subset of the total features. Training a model on a bootstrapped training set is also called *bagging* in literature, and this technique, although somewhat efficient, fails to deliver highly accurate models owing to the high correlations the resulting trees would exhibit. The additional constraint on the random set of features the model is allowed to choose for each split reduces the correlation between trees, thus allowing further variance reduction in the random forest ensemble.

The tree-boosting algorithm, instead of bagging and randomizing the feature set to select for each split, fits *shallow* trees (trees of limited depth) sequentially on the residuals of target and prediction. Since the model is exposed to the residuals multiple times, large errors on the outliers can be more easily handled, and the sequential learning only gradually increases the variance, thus making the model less prone to overfitting (Witten et al., 2021). The algorithm, described in detail in Witten et al. (2021), is here included in Algorithm 1 for ease of reference. The parameter  $\lambda$  allows us to modulate how quickly the algorithm learns. Smaller values of  $\lambda$  guarantee a more robust training which usually results in lower out-of-sample error. However, a trade-off between  $\lambda$  and the number of trees  $B$  must be operated, as a smaller  $\lambda$  will in general require a higher number of trees to learn the data appropriately.

LightGBM is a fast implementation of the tree-boosting logic since the continuous predictors are discretized and binned, thus generating a histogram. Such histograms of the predictors are then used to determine the split points which in turn establish the feature space partition.

In the context of this project, the hyperparameters of the LightGBM model are selected via 3-fold cross-validation. In addition, LightGBM can only handle one-dimensional targets, therefore we need to build 16 distinct models to operate the forecast (16-week-long targets, equal to the test set size).

Finally, LightGBM - like most gradient boosting libraries - does not automatically output the prediction intervals, thus we will operate quantile regression using the quantiles 0.025, 0.5, and 0.975 (where the extrema constitute the 95% prediction interval and the 0.5 quantile - the median - is to be interpreted as forecast). The reader should note how, with the inclusion of the prediction intervals, a total of 48 models (3 quantiles  $\times$  16-week long targets) need to be trained on the data. However, owing to its efficient implementation, LightGBM is rather fast even without resorting to hardware accelerators such as GPUs.

**Result:** Boosting for Regression Trees

Set  $\hat{f}(x) = 0$  and  $r_i = y_i$  for all  $i$  in the training set  
**for**  $b = 1$  to  $B$  **do**  
  Fit a tree  $\hat{f}^b$  with  $d$  splits ( $d + 1$  terminal nodes) to the training data  $(X, r)$   
  Update  $\hat{f}$  by adding in a shrunken version of the new tree:  

$$\hat{f}(x) \leftarrow \hat{f}(x) + \lambda \hat{f}^b(x)$$

Update the residuals:  $r_i \leftarrow r_i - \lambda \hat{f}^b(x_i)$

Output the boosted model:

$$\hat{f}(x) = \sum_{b=1}^B \lambda \hat{f}^b(x)$$

**Algorithm 1:** Boosting for Regression Trees, from [Witten et al. \(2021\)](#).

### 3.3.4. LSTM

Long-Short Term Memory Networks (LSTMs) ([Hochreiter and Schmidhuber, 1997](#)) are a subclass of Recurrent Neural Networks (RNNs) ([Rumelhart et al., 1986](#)) whose aim is to solve the issues related to the inability of the simple RNNs to capture long-term dependencies. LSTMs are best explained through visualization, which we included in Figure 3.2. Given an input  $x_t \in \mathbb{R}^d$  and a hidden state for the previous time step  $h_{t-1} \in \mathbb{R}^c$ , a recurrent unit:

- vertically concatenates the input  $x_t$  and  $h_{t-1}$ :

$$y_t = [x_t; h_{t-1}] \in \mathbb{R}^{c+d}$$

- operates a matrix multiplication (by matrix  $W \in \mathbb{R}^{c \times (c+d)}$ ) and a bias addition (by vector  $b \in \mathbb{R}^c$ ):

$$z_t = W y_t + b \in \mathbb{R}^c$$

- applies a non-linear transformation (an activation function, typically a sigmoid or a hyperbolic tangent):

$$h_t = \sigma(z_t) \in \mathbb{R}^c$$

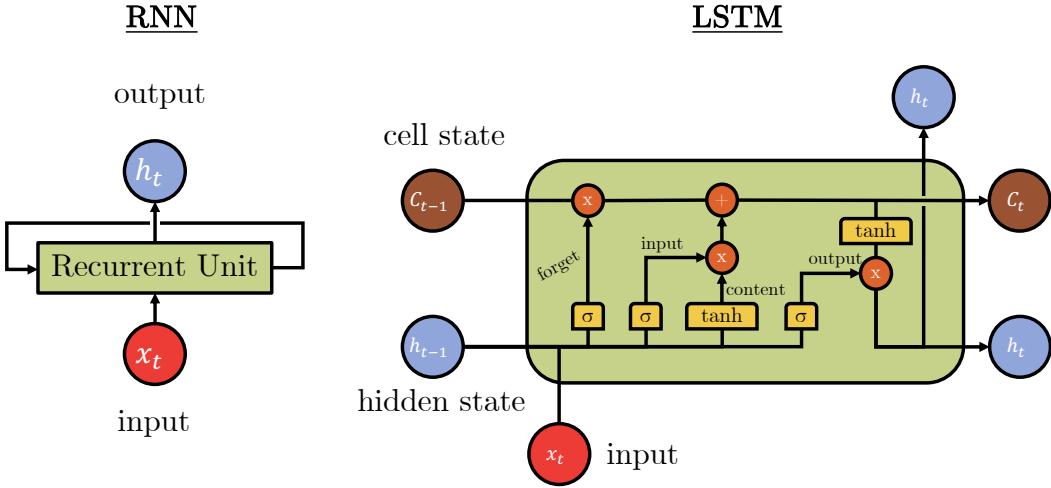
where, for univariate time series forecasting,  $c$  and  $d$  are both equal to 1. The power of RNNs lies in the weight sharing this architecture offers, since the matrix  $W$  and the bias  $b$  are constant across all time steps (although they are trainable parameters).

LSTM cells extend the RNN architecture using a cell state  $c_t$ , which acts as cell memory, and 4 gates - establishing how the information is retained/removed by the cell - are introduced, each with their own weight matrices and biases. The tasks of each of the 4 gates, along with the equations, to be applied in reference to Figure 3.2, are included below.

- The *forget gate*,  $f_t$ , which determines how much information is to be removed from the cell state:

$$f_t = \sigma(W_f[x_t; h_{t-1}] + b_f)$$

- The *input gate*,  $i_t$ , and the *content gate*,  $k_t$ , which define whether and where to write



**Figure 3.2.:** *Left:* scheme of a recurrent unit. *Right:* an LSTM cell. Inspired by [Colah \(2015\)](#).

new information in the cell state:

$$i_t = \sigma(W_i[x_t; h_{t-1}] + b_i)$$

$$k_t = \tanh(W_c[x_t; h_{t-1}] + b_c)$$

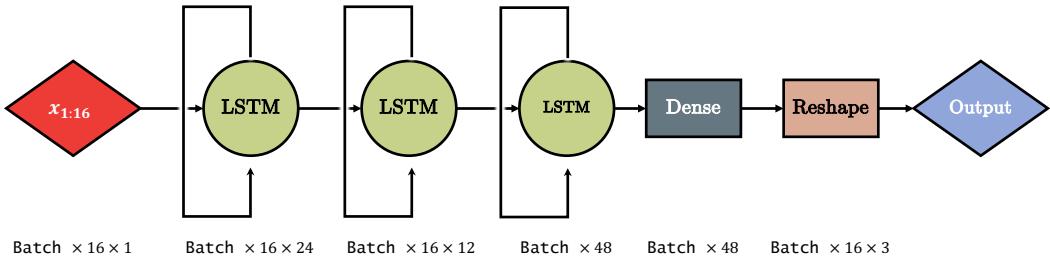
- The *output gate*,  $o_t$ , which decides how much of the cell state is to be included in the output:

$$o_t = \sigma(W_o[x_t; h_{t-1}] + b_o)$$

The decision on the LSTM architecture is made, in this research, by resorting to [Keras Tuner](#) and using, as metric, the cross-validation on the entire training set across all the time series. Given computational constraints and a possibly unlimited amount of combinations, we have decided against the selection of a specific architecture for each time series. The chosen LSTM architecture can be found in Figure 3.3. The pinball loss is used for training, with quantiles 0.025, 0.5, and 0.975. Adam is chosen as the optimizer for the training phase. In addition, we set a maximum number of 1000 epochs, although we include an early-stop callback to interrupt the training and restore the best weights if the validation pinball loss does not decrease for 10 consecutive epochs. Please refer to the [associated GitHub repository](#) for the complete code and layer shapes.

### 3.3.5. CNN

Convolutional Neural Networks (CNNs) are inspired by the way the human brain processes images, and popularized by LeCun's pioneering work in the field of artificial intelligence ([LeCun et al. \(1989\)](#) and [LeCun et al. \(1998\)](#)). Although the initial use of CNNs was limited to computer vision, more recent advancements ([Gehring et al., 2017](#)) have paved the way for their application in sequence-to-sequence translation tasks as well, which intuitively justifies their deployment in the realm of time series forecasting. We include a visualization of a 1D-Convolution on a time series in Figure 3.4. Let  $\{X_t\}$  be a time series of length  $T$  to which a  $K$ -long 1D kernel  $k$  is applied, the resulting time series  $\{Y_t\}$ , in the simplest scenario, will have output size  $T - K + 1$  and each of the resulting entries can be written in the following

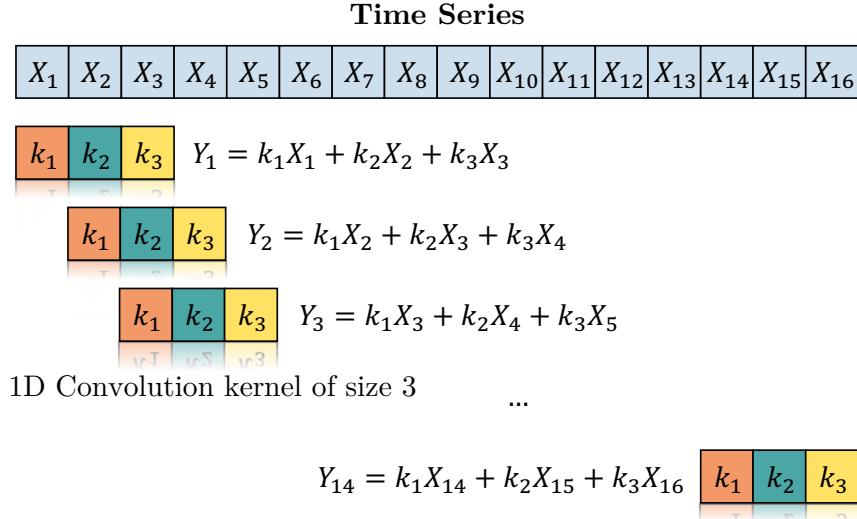


**Figure 3.3.:** The chosen LSTM architecture. Batch indicates the batch dimension (13 for time series-wise models and 128 for data set-wise models). The number 3 in the output represents the quantiles 0.025, 0.5, 0.975 for 95% prediction interval and median forecast. The three LSTM layers feature, respectively, 24, 12, and 48 units, whereas the fully-connected Dense layer contains 48 neurons.

way.

$$Y_t = \sum_{j=1}^K k_j X_{t+j-1} \quad \text{for } t \in \{1, \dots, T - K + 1\}$$

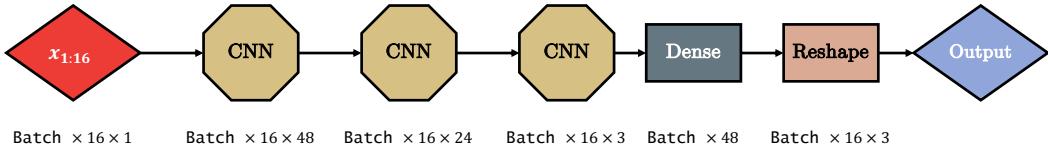
Once again, the choice of architecture for the CNN has been automated through the use of Keras Tuner. The final architecture for the network can be found in Figure 3.5. The training procedure is identical to the one adopted for the LSTM. Please refer to Section 3.3.4 for the details.



**Figure 3.4.:** Convolution of a 16-step long time series with a 1D kernel of size 3.

### 3.3.6. Gaussian Processes

Gaussian Processes (GPs) are Bayesian non-parametric models which effectively represent a distribution over a set of functions. We only provide a high-level description of the model in connection with the current project - which mainly focuses on automatic kernel selection - but more detailed information is provided in [Rasmussen and Williams \(2005\)](#) and [Heard \(2021\)](#). As we are focusing on time series, the following description will be based on the time



**Figure 3.5.:** The chosen CNN architecture. Batch indicates the batch dimension (13 for time series-wise models and 128 for data set-wise models). The number 3 in the output represents the quantiles 0.025, 0.5, 0.975 for 95% prediction interval and median forecast. The Convolution1D layers - in the image named CNN - feature, in order, 48, 24, and 3 kernels each of size 3, with same padding. The final fully-connected Dense layer contains 48 neurons. ReLU activations are used across the network.

series notation introduced so far. Note that, when using GPs for time series, we instantiate a vector indicating the time steps (the time does not necessarily be physical; non-dimensional time is acceptable as well).

**Definition 3.3.2** (*Gaussian Process*, adapted from Heard (2021)). Let  $\{X_t | t \in T\}$  be a time series,  $t \in \mathbf{t} = [t_1, t_2, \dots, t_n]$  be the vector of the time predictor,  $m$  be the *mean function* such that  $m : T \rightarrow \mathbb{R}$  and  $k$  be the *covariance* or *kernel function* such that  $k : T \times T \rightarrow \mathbb{R}$ , then  $\{f(t) | t \in T\}$  is a *Gaussian Process* and we commonly write  $f(t) \sim \text{GP}(m, k)$  if, for all  $t \in \mathbf{t}$ :

$$f(\mathbf{t}) \sim \mathcal{N}(m(\mathbf{t}), K(\mathbf{t}, \mathbf{t}))$$

and

$$K(\mathbf{t}, \mathbf{t})_{ij} = [k(t_i, t_j)]$$

The covariance function must be symmetric and positive semidefinite, so that, for all  $t \in \mathbf{t}$  and  $a_i, a_j \in \mathbb{R}$ , the inequality  $\sum_{i=1}^n \sum_{j=1}^n a_i a_j k(t_i, t_j) \geq 0$  holds.

In most of the real-world scenarios, we do not have access to the noiseless function  $f(t_i)$ , but rather the data is collected from the realization of the underlying Gaussian Process with the inclusion of some noise  $\epsilon_i$ , so that  $y_i = f(t_i) + \epsilon_i$ . If the noise is Gaussian, then  $\epsilon_i \sim \mathcal{N}(0, \sigma^2)$ , then  $y_i \sim \mathcal{N}(f(t_i), \sigma^2)$  and, as explained in Heard (2021) Proposition 10.1, the posterior distribution is again another Gaussian Process where:

$$f | \mathbf{t}, \mathbf{y} \sim \text{GP}(m^*, k^*)$$

and the posterior mean and covariance functions are:

$$m^*(t_i) = m(t_i) + k(t_i, \mathbf{t}) \{K(\mathbf{t}, \mathbf{t}) + \sigma^2 I\}^{-1} (\mathbf{y} - m(\mathbf{t})) \quad (3.4)$$

$$k^*(t_i, t_j) = k(t_i, t_j) - k(t_i, \mathbf{t}) \{K(\mathbf{t}, \mathbf{t}) + \sigma^2 I\}^{-1} k(\mathbf{t}, t_j). \quad (3.5)$$

The two equations above specify what the form of the Gaussian Process is *after* observing the data, but no direct information about the kernel's parameters can be inferred from it. In order to set such hyperparameters, one could either leverage prior knowledge - if available - or a data-driven method may be defined. As we specified at the beginning of this project, our aim is to make the models accessible to everyone, regardless of their background. Therefore, a data-driven approach would be more appropriate to fulfill this objective. Among the data-driven approaches, we have two options: full Bayesian inference via Markov Chain Monte Carlo (MCMC), or the search of point estimates for the hyperparameters via Maximum Likelihood Estimation (MLE). We choose the latter to limit the computational burden.

Equation 2.30 from [Rasmussen and Williams \(2005\)](#) shows that the marginal log-likelihood of the Gaussian Process can be written as:

$$\log p(\mathbf{y}|\mathbf{t}) = -\frac{1}{2}(\mathbf{y} - \mathbf{m}(\mathbf{t}))^\top(K(\mathbf{t}, \mathbf{t}) + \sigma^2 I)^{-1}(\mathbf{y} - \mathbf{m}(\mathbf{t})) - \frac{1}{2}\log|K(\mathbf{t}, \mathbf{t}) + \sigma^2 I| - \frac{n}{2}\log 2\pi$$

This equation can be used to find the appropriate kernel hyperparameters and entails the minimization of the negative log-likelihood of the Gaussian Process. In this project, we use the Nelder-Mead optimizer ([Nelder and Mead, 1965](#)) to minimize the negative log-likelihood.

The automatic construction of the kernels for the GP is the last topic to be discussed. Many of the ideas are derived from [Kronberger and Kommenda \(2013\)](#), [Duvenaud et al. \(2013\)](#), [Lloyd et al. \(2014\)](#), and [Duvenaud \(2014\)](#), although we introduce the novelty in the kernel selection process, moving away from the BIC (see Definition 3.1.1) and instead adopting the WBIC (see Definition 3.1.2) which is non-tractable in our case, and we, therefore, employ the Monte Carlo estimate we proposed for it (see Equation 3.3).

In order to explore the space of possible kernels we need three components:

- the definition of a set of base kernels.
- the definition of the kernel combination operations.
- the definition of a metric for model selection.

We will explain these three items in detail.

### Base Kernels

The definition of base kernels is itself a matter of preference. [Lloyd et al. \(2014\)](#) included only the 6 distinct base kernels, namely: 1) white noise kernel; 2) constant kernel; 3) squared exponential kernel; 4) periodic kernel; 5) linear kernel; 6) polynomial kernel. [Kronberger and Kommenda \(2013\)](#) also included the Matérn kernel. We find that the inclusion of the Matérn kernel tends to lead to models which are less interpretable and follow closely the noise in the training set, and we decided to exclude it instead. In addition, the inclusion of the polynomial kernel in the context of a univariate input is pointless, since the product of linear kernels would automatically give rise to a polynomial kernel, should it be needed to explain the data. Thus the polynomial kernel is excluded from the base set as well, although we will describe it for reasons of completeness. A brief overview of such set follows and a visualization of the prior samples from the 6 base kernels used in [Lloyd et al. \(2014\)](#) is included in Figure 3.6. The interested reader should refer to [Rasmussen and Williams \(2005\)](#) for further details.

Let  $t, t' \in T$  where  $T$  is the time set under study.

**Definition 3.3.3 (White Noise Kernel).** The white noise kernel models the uncorrelated noise in the time series. The covariance function representing the kernel is given by:

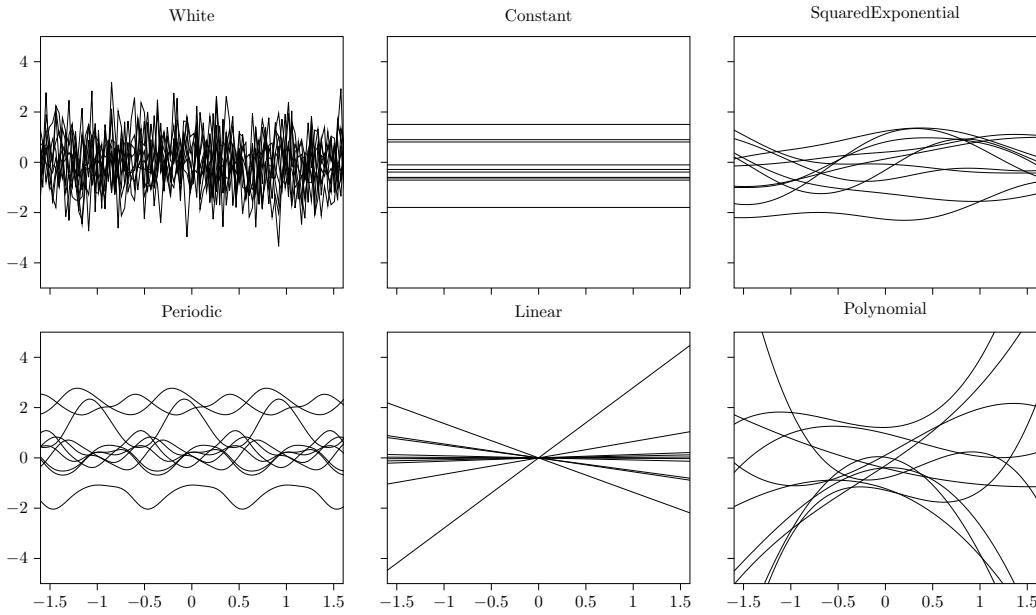
$$k(t, t') = c\mathbb{1}_{\{t=t'\}}$$

where  $c \in \mathbb{R}$  is an appropriate constant (learned from the data).

**Definition 3.3.4 (Constant Kernel).** The constant kernel is used to add an offset to the underlying function the GP needs to learn. The corresponding covariance function is a simple constant for all data points:

$$k(t, t') = c$$

where  $c \in \mathbb{R}$  is a constant.



**Figure 3.6.:** 10 random samples from the prior for each of the 6 kernels used as a base set in Lloyd et al. (2014). We opted for the exclusion of the polynomial kernel in this study.

**Definition 3.3.5 (Squared Exponential Kernel).** The exponential kernel is the default and is used in most circumstances owing to its flexibility. Its covariance function can be written as:

$$k(t, t') = \exp \left\{ -\frac{(t - t')^2}{2l^2} \right\}$$

where the constant  $l$ , also referred to as *length scale*, scales the distance and modulates how quickly the mutual influence of two points decays.

**Definition 3.3.6 (Periodic Kernel).** The periodic kernel is utilized whenever seasonal components are observed in the data. The kernel can be expressed as:

$$k(t, t') = \exp \left\{ -\frac{2}{l^2} \sin^2 \left( \pi \frac{|t - t'|}{p} \right) \right\}$$

where  $l \in \mathbb{R}$  is a length scale, with exactly the same meaning as in the Squared Exponential Kernel, and the parameter  $p$  can be interpreted as a *period*, modulating the frequency of the resulting functions. A sine function is introduced to model the periodic nature of the covariance.

**Definition 3.3.7 (Linear Kernel).** The linear kernel (also commonly known as *Dot Product Kernel*) can be shown to be the kernel associated with performing linear regression through a Gaussian Process. If the offset ( $c$  in the following equation) is zero, then the kernel is said to be *homogeneous* (which is equivalent to a linear regression model without the intercept).

$$k(t, t') = c + att'$$

where  $c, a \in \mathbb{R}$  are two model's constants.

**Definition 3.3.8 (Polynomial Kernel).** The polynomial kernel is a generalization of the

linear kernel, where the kernel is now raised to a power  $d$ , the degree of the polynomial chosen to be a candidate for the function's fit.

$$k(t, t') = (c + att')^d$$

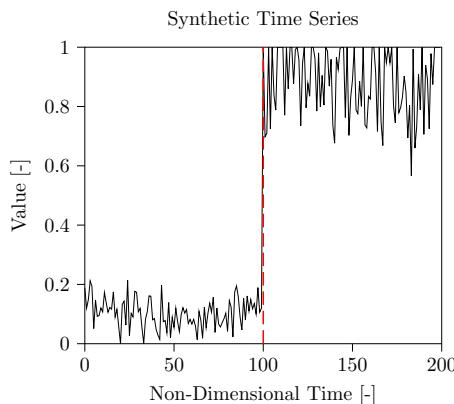
Where  $c, a, d \in \mathbb{R}$  are constants learned from data. Optimizing polynomial kernels is a slow process. One might argue that if the data truly originated from a GP with a polynomial kernel, then an automatic kernel selection program should reconstruct it using a product of linear kernels. This rationale supports their exclusion from the study.

### Kernel Combination Operations

First, we recall that sum and product of two kernels result in a valid kernel. The reasoning behind such a claim is provided below.

- **Sum:** kernels are positive semidefinite functions, resulting in positive semidefinite matrices. Consider a kernel  $k_1$  associated with a covariance matrix  $A \in \mathbb{R}^{n \times n}$  and a kernel  $k_2$  associated with a covariance matrix  $B \in \mathbb{R}^{n \times n}$ , it must be that if  $k_3 = k_1 + k_2$ , then the matrix  $C \in \mathbb{R}^{n \times n}$  associated with  $k_3$  is  $C = A + B$  and that  $x^T C x = x^T A x + x^T B x \geq 0$  since both  $x^T A x \geq 0$  and  $x^T B x \geq 0$  for all  $x \in \mathbb{R}^n$  because of their positive definiteness. This implies that  $C$  is positive semidefinite and symmetric (since  $A$  and  $B$  are covariance matrices and therefore symmetric). Thus the sum of two valid kernels is again a valid kernel.
- **Product:** kernels are associated with covariance matrices that are symmetric and positive semidefinite by definition. The product of two kernels is to be interpreted, in a matrix context, as the Hadamard product of the associated covariance matrices. The Schur Product Theorem states that the Hadamard product of two symmetric positive semidefinite matrices is symmetric and positive semidefinite, which leads to the conclusion that the product of two kernels is itself a valid kernel.

The extra operation we include is the change point, which can be defined as a point in time where the behavior of the time series drastically changes. Change points cannot be *forecast*. However, if they are detected in the history of a time series, they allow to neglect portions of the data that are not deemed particularly useful for the forecast.



**Figure 3.7.:** Example of a change point. The dashed red line indicates the change point location. A sudden change in mean and variance can be observed at approximately  $t = 100$ .

The introduction of the change point in the kernel construction may complicate the exploration of the kernel space. Some restrictions we place on the change point (similar to those included in [Lloyd et al. \(2014\)](#)) may simplify the problem. These will be made clear in Algorithm 2 (where we will refer to the change point as  $CP$ ). Note how the change point can only replicate the best kernel found until a given iteration, or simply prepend/append a constant kernel: this simplification is operated in order to reduce the size of the exploration space.

**Result:** Kernel Discovery

Set base kernels  $K = \{k_{\text{white}}(t, t'), k_{\text{constant}}(t, t'), k_{\text{sq.exp.}}(t, t'), k_{\text{periodic}}(t, t'), k_{\text{linear}}(t, t')\}$

Initialize best kernel  $k^*$

**while** *The evaluation metric for the best kernel  $k^*$  improves* **do**

Retrieve best kernel up to the given iteration:  $k^*$  (initialize at first iteration)

**for**  $k' \in K$  **do**

$k = k^* + k'$  // perform kernel sum

Fit GP using  $k$  as kernel

Calculate and store the evaluation metric

**for**  $k' \in K$  **do**

$k = k^* \times k'$  // perform kernel product

Fit GP using  $k$  as kernel

Calculate and store the evaluation metric

**for**  $k' \in \{CP(k^*, k^*), CP(k^*, k_{\text{constant}}), CP(k_{\text{constant}}, k^*)\}$  **do**

$k = k'$  // apply change point kernel

Fit GP using  $k$  as kernel

Calculate and store the evaluation metric

Select the kernel  $k^*$  associated to the model with the best evaluation metric at the given iteration.

**Algorithm 2:** Automatic kernel discovery algorithm.

### Metric for model selection

In Algorithm 2, we referred to a vague *evaluation metric* to be used for model selection. Such metric is chosen to be, in [Duvenaud et al. \(2013\)](#), [Lloyd et al. \(2014\)](#), and [Duvenaud \(2014\)](#), the BIC. For the reasons explained in Section 3.1.1, the BIC is unsuitable as an evaluation metric in the context of GPs, and instead, we proposed a Monte Carlo approximation for the WBIC 3.1.2 which is instead more appropriate for nonparametric models. The prior for the model parameters (all positive by construction) is set to be  $\text{Gamma}(1, 1)$ . Within the scope of this project, we will attempt model selection with both BIC and WBIC and compare the obtained results, in order to provide support to our claims of superior performance of the WBIC-based method. Other options, such as cross-validation on the training set, may also be valid, although the computational cost of the GP fitting increases as a cubic function of the number of points, thus making the use of cross-validation a rather unappealing alternative.

#### 3.3.7. Hybrid Methods

The category of Hybrid Methods encompasses any model combining two or more of the previously described methods to generate a forecast. As such, there are no rigid specifications concerning the scope of these models or the architecture they must adhere to. For reasons of computational limitations, we will focus on a specific type of hybrid models, featuring a local-global scope. Note that using hybrid methods in this circumstance leverages the knowledge the model obtains from multiple time series, therefore the data set is assumed to contain more than a single time series. This may or may not be acceptable, depending

on the application, but when we perform search engine optimization, we usually have access to a large set of keywords, thus using hybrid methods would not be an issue. The three proposed models are named in the following way:

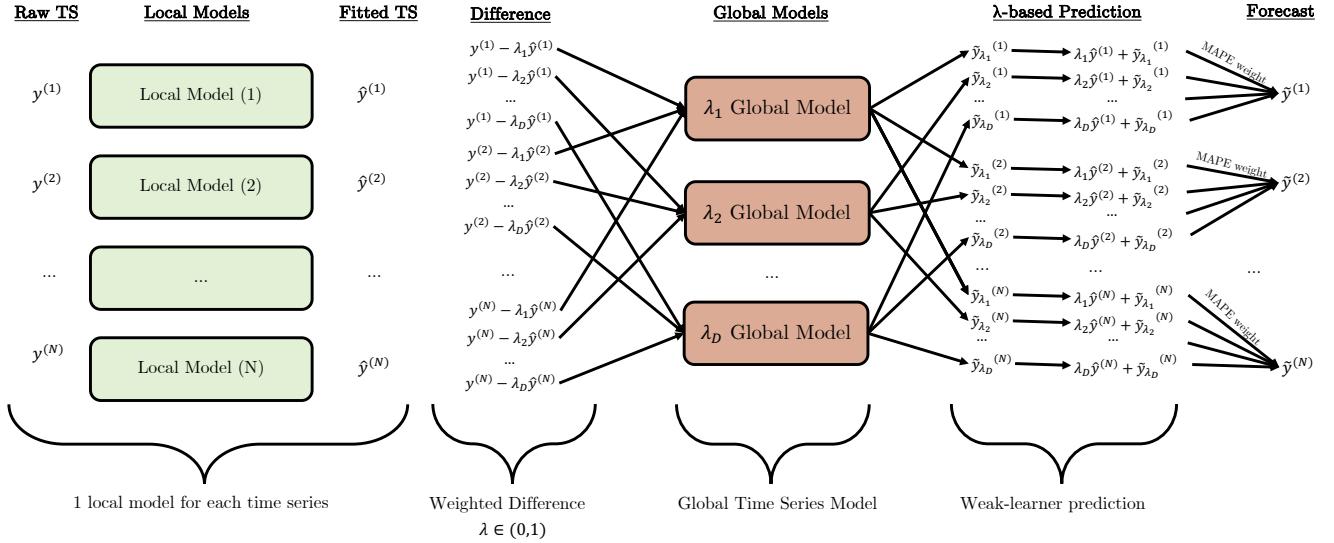
- **Weak Learner + Global Model (WLGM)**: in this model, each of the time series in the data set is fitted using a GP with automatic covariance detection, and the in-sample prediction is obtained. Afterward, the in-sample prediction, modulated by a scalar  $\lambda \in [0, 1]$  - which is why we refer to such a model as a weak learner - is subtracted from the original time series to obtain a residual. A model is then trained on the residuals for each value of the chosen  $\lambda$  (in the context of this research, we chose 10 values of  $\lambda$  equally spaced between 0.5 and 1.0). The output obtained from each of the time series for these global models is then summed to the weak learners' prediction and weighted by the normalized in-sample MAPE. For the out-of-sample forecast, we follow exactly the exact same logic except for the MAPE (which relies on the training set estimates, and not the test set ones, otherwise data leakage would occur).
- **Weak Learner + Clustering Model (WLCM)**: this works in the same way as WLGM, but instead of training global models on the  $\lambda$ -weighted residuals, clustering is performed on the training set of the original time series, thus allowing a different model for each of the cluster at a given  $\lambda$ . Clustering on the time series is performed using Dynamic Time Warping (see Definition 2.2.13) as distance metric and HDBSCAN ([Campello et al., 2015](#)) ([McInnes et al., 2017](#)) to execute the actual clustering - although other options for the clustering algorithm are also possible - where for a new cluster to be created we impose that this must contain at least 5 distinct time series.
- **Weak Learner + Residual Clustering Model (WLRCM)**: WLRCM resembles WLCM with the only difference that the clusters are calculated on the residual  $\lambda$ -weighted time series instead of the original ones.

Figures 3.8, 3.9, and 3.10 are self-explanatory, but we specify the extra information needed to reproduce our results in detail: the local model is chosen to be a Gaussian Process trained via automatic kernel discovery with WBIC, the global models or cluster wise models feature the same architecture as the CNNs described in Section 3.3.5, and 10 distinct values for  $\lambda$  are set to be equally spaced in the interval [0.5, 1.0].

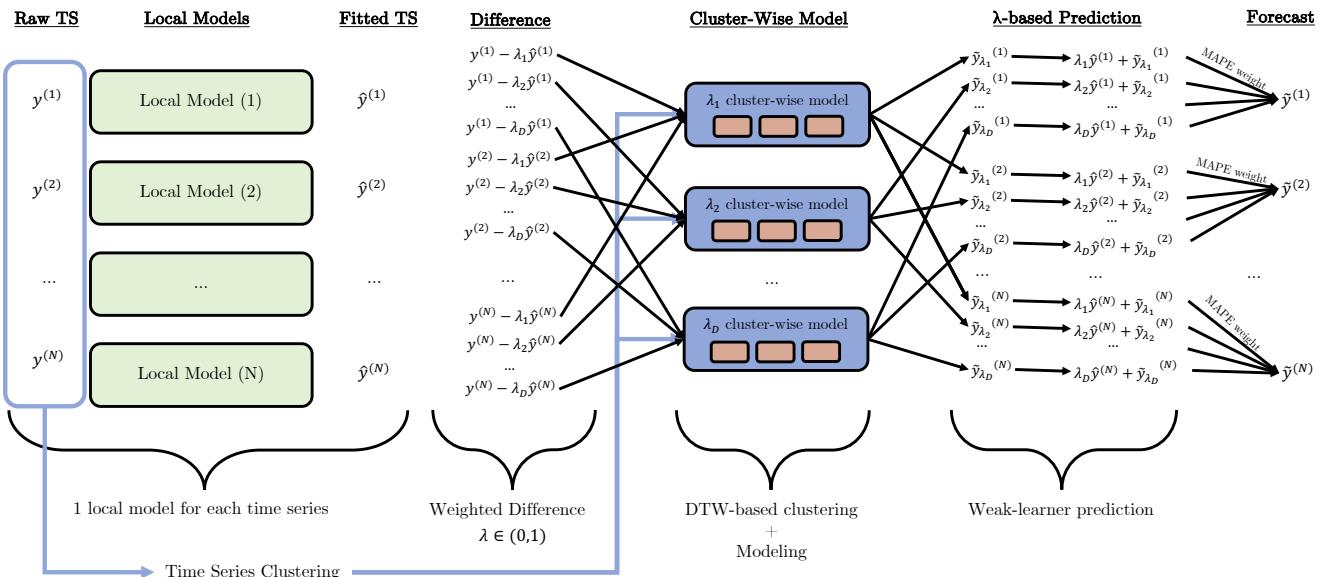
The reason behind the potential improvement the hybrid methods we propose may offer is related to two distinct properties:

- the models include the parameter  $\lambda$  which generates different data sets for each value of such parameter, thus offering similar properties to bagging, but with reduced correlation.
- the errors the models make may complement each other, thus the ensemble may offer more stable and reliable predictions. Furthermore, the use of ensembles in the hybrid models may reduce the variance of the prediction and, as an implication, the width of the prediction intervals (please note that the prediction interval width estimation for these hybrid models will not be derived in the current work, but we will only describe the results in terms of the mean prediction).

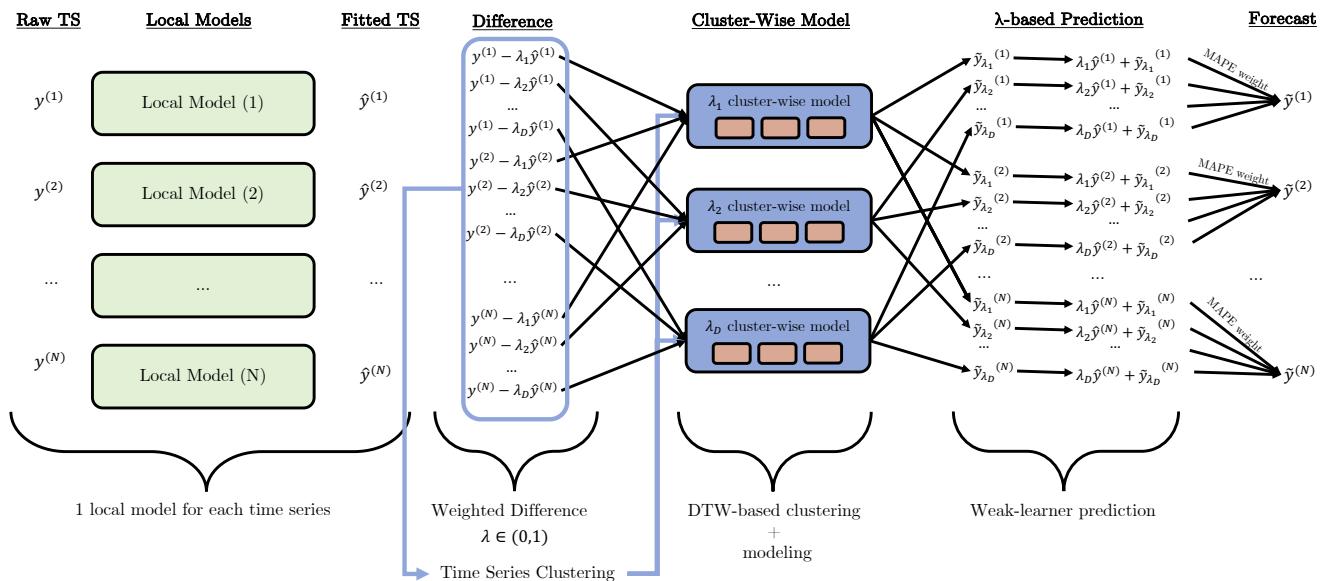
This completes the section related to the employed methods for this study.



**Figure 3.8.: Weak Learner + Global Model (WLGM):** This hybrid method uses a local model to predict the main characteristics of the time series. The model's prediction is then *weakened* via the multiplication with a scalar  $\lambda$  and global  $\lambda$ -wise models are fitted before the predictions are recombined via MAPE weighting.



**Figure 3.9.: Weak Learner + Clustering Model (WLCM):** Prediction based on weak learner residuals and original time series clustering. The  $\lambda$ -wise forecast is weighted via training set MAPE.



**Figure 3.10.: Weak Learner + Residual Clustering Model (WLRCM):** Prediction based on weak learner residuals and clustering of such residuals. The  $\lambda$ -wise forecast is weighted via training set MAPE.

## 4. Pre-processing

This chapter will describe how the data set used in this project has been retrieved from the available online resources and how it has been pre-processed before feeding it to the models described in Chapter 3.

### 4.1. Data Retrieval

The initial challenge we need to address concerns the retrieval of the data from online resources. According to [Statista \(2023\)](#), Google detains approximately 85% of the worldwide market share, making it the primary go-to resource for online searches.

[Google Trends](#) is an online platform maintained by Google which keeps track of the keyword search trends on the search engine. In particular, Google Trends attributes a value to a keyword (which can contain multiple words) ranging from 0 to 100, where 0 defines a word that has never been searched during the time interval of interest and 100 is used to indicate a word that has been searched as many times as the most searched keyword on Google. This means we do not have direct access to the raw counts for the keywords' searches, but we have access to a proxy for such value, as the relative ordering of the keywords is preserved and thus comparisons between different keywords remain meaningful.

The problem, therefore, shifts from *obtaining the data* to *selecting the keywords* whose search time series can be obtained from Google Trends.

For this purpose, we resort to three distinct online tools:

- [Google's Machine Learning Glossary](#)
- [OpenAI's ChatGPT-4](#)
- [Web of Science](#)

Each of the three resources above caters to a different group of users:

- Google's Machine Learning Glossary targets a broad variety of individuals, ranging from beginners to experienced practitioners who want to learn or refresh their knowledge about some of the main topics in Machine Learning. However, the fundamental issue with the glossary is the lack of non-Google-related products or terms. As an example, PyTorch - a product of Facebook AI and the most widely used library in the Deep Learning community - is absent, whereas TensorFlow (its direct Google competitor) is included.
- OpenAI's ChatGPT-4 is a Large Language Model (LLM) which has been trained on [Common Crawl](#), an open repository of web crawl data. Books, news articles, academic journals, and other materials available on the web are included in the data set, triggering a broad discussion regarding privacy concerns, the role of the individual in a society where AI can accomplish complex tasks, and fair use of training data. The model is therefore exposed to a multitude of machine learning and artificial intelligence articles, making it rather suitable to answer questions about trends and popular keywords in the context of AI.

- Web of Science is a scientific citation indexing service that is highly popular in academia. The strength of this service lies in the possibility to search for *hot papers* - a selection of journal articles that have been cited many more times than their counterparts in the same field released in approximately the same time interval - with the relative describing keywords.

We will discuss the keyword lists obtained by each of the three sources.

#### 4.1.1. Google’s Machine Learning Glossary

Google’s Machine Learning Glossary is distributed under the Creative Commons Attribution 4.0 License, thereby allowing distribution and modification as long as credit to Google is provided. We parsed all the keywords included in the glossary and created a word cloud that allows quick visualization of the keyword list. When a keyword contained multiple words, we connected the different words with an *underscore* “\_”. The result is included in Figure 4.1. Please note that the complete list is not included in the chapter body in order to reduce its size. The full list of the 538 keywords is provided in the relative section in the appendix.



**Figure 4.1.:** Word cloud of the Google Machine Learning Glossary

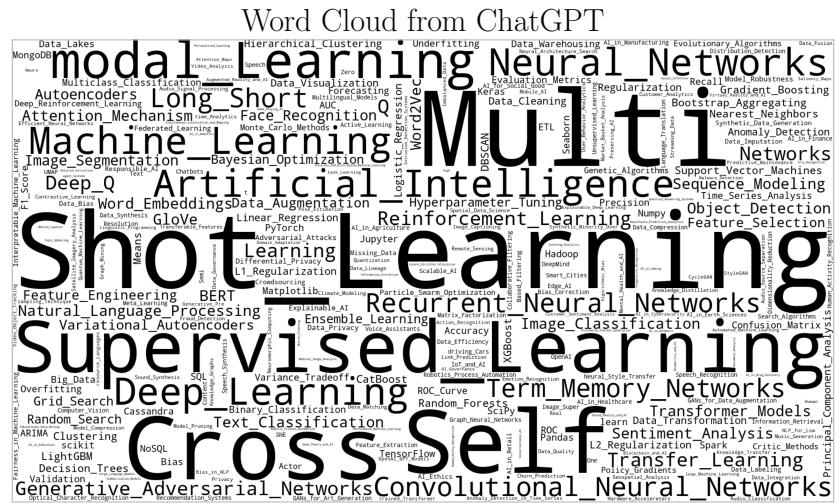
The size of the keywords does not bear a particular meaning as all keywords appear only once in the data set. In the word cloud, we can recognize a multitude of Machine Learning-related concepts, such as Shot Learning, Imbalanced Dataset, Area Under the ROC Curve, and so on. Looking a little longer allows us to spot words such as Keras and Tf, where the former is the high-level wrapper for Google's Deep Learning library Tensorflow and the latter is an alias commonly used to indicate the Tensorflow library itself.

#### 4.1.2. OpenAI's ChatGPT-4

Next, we asked ChatGPT-4 to provide a list of AI-related keywords using the following prompt:

Hello ChatGPT! I am looking for a list of approximately 500 keywords of the hottest research and development topics in the fields of Machine Learning, Deep Learning, and Artificial Intelligence. Please provide the list as a Python list structure.

ChatGPT-4 returned a total of 283 keywords, which can be found in the appendix. These are also provided, as before, in the word cloud included in Figure 4.2. Some of the keywords



**Figure 4.2.:** Word cloud of the AI-related keywords from ChatGPT-4

appear in both the Google Machine Learning Glossary and the list provided by ChatGPT-4, and in some instances, the words appear in singular/plural forms.

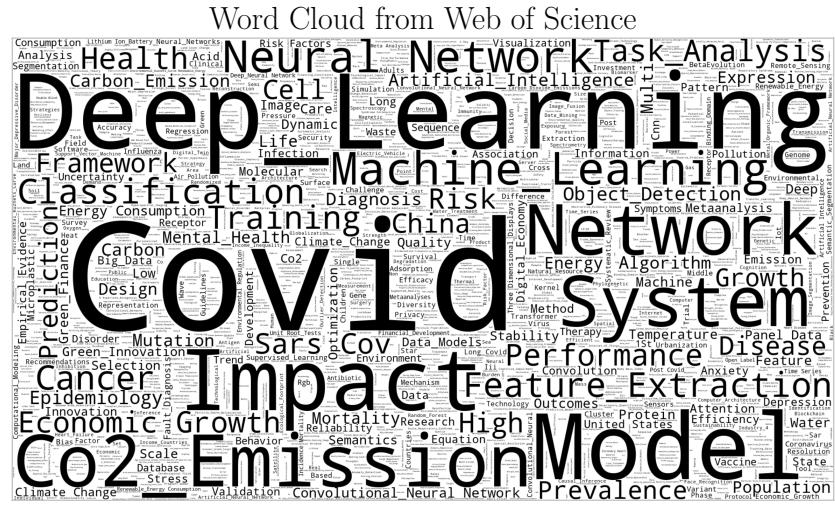
#### 4.1.3. Web of Science

On Web of Science, we queried all the scientific articles relating to either Machine Learning, Deep Learning, or Artificial Intelligence. The number of articles was in the order of millions, thus we filtered only the *hot papers*. The query retrieved a total of 1081 papers, of which we then obtained the describing keywords. Once again, the full list of the keywords is provided, in alphabetical order, in the appendix. A word cloud of the keywords has also been produced, and it can be found in Figure 4.3. We notice that many of the keywords included in this word cloud concern the topics of green energy, pollution, COVID-19, and healthcare. These keywords do not look specific enough to the fields of Machine Learning, Deep Learning, and Artificial Intelligence, therefore we decided not to include the Web of Science keywords in this study. Please note that we do believe that Web of Science is a reputable source with a large impact in academia, but we deem it unsuitable *for the purpose of this research*.

#### 4.1.4. List Merging and Data Set Creation

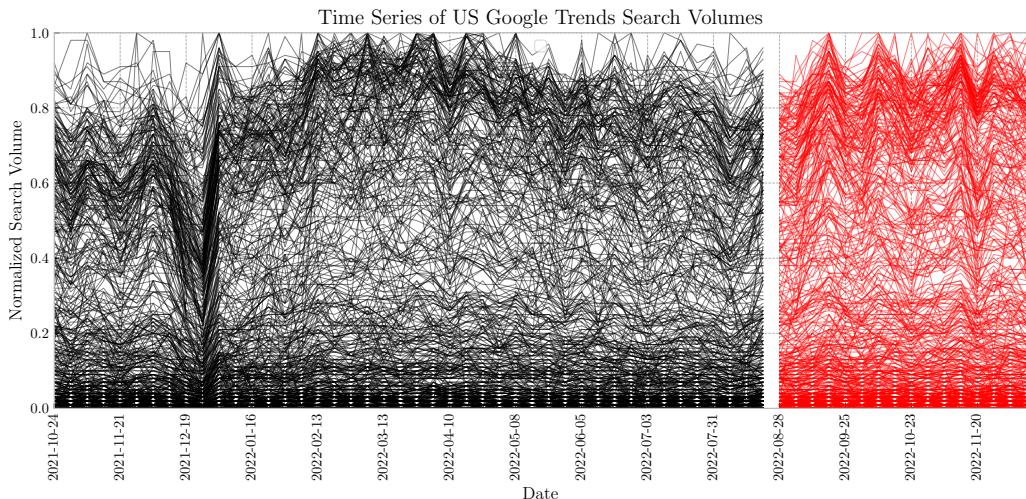
The keyword lists from the Google Machine Learning Glossary and from ChatGPT-4 have been inflected to their singular form before being merged. This led to the creation of a keyword set of a total of 802 keywords.

We then proceeded with querying US Google Trends for the weekly time series of the search volumes of each of the keywords in the list, in the time range going from the 24th



**Figure 4.3.:** Word cloud of the keywords obtained from the hot papers on Web of Science for the topics of Machine Learning, Deep Learning, and Artificial Intelligence.

of October 2021 to the 11th of December 2022 (60 weeks in total <sup>1</sup>). We then split the training set and the test set, and the output of this procedure can be found in Figure 4.4. In particular, as a test set, we consider a wide time window of 16 weeks, which will allow us to judge both the short-term and long-term forecasts of the considered techniques, leaving us with 44 weeks for the training set.



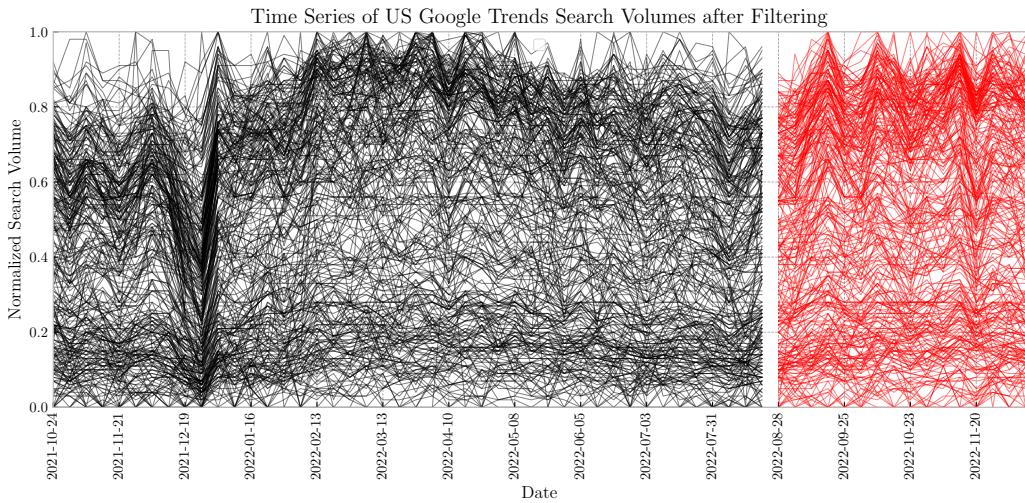
**Figure 4.4.:** Time Series plot for the 802 keywords in the data set. The black portion of the lines indicates the training set, whereas the red portion of the lines defines the test set.

<sup>1</sup>Please note that the 60-week value is chosen so that slightly more than one year is covered. Longer time series could be considered but we decided against this option given 1) computational constraints and 2) the fact that AI is a field where innovation happens at a fast pace, therefore new topics of interest may quickly appear, leaving the analyst without much time history to reason on.

## 4.2. Exploratory Data Analysis, Cleanup, and Visualization

Figure 4.4 shows that, as expected, the time series are bounded in an interval of 0 to 100. The reader will not fail to notice that a large number of time series in Figure 4.4 are very close to zero and that some of them do not seem to vary considerably across the considered time frame. This is an issue, as the data set would tend to be imbalanced in favor of series featuring small values. In addition, when considering some keywords for SEO, we are usually interested in predicting the increase in popularity of a given keyword - and thus the trend - instead of whether the popularity of a given keyword will remain constant.

With these ideas in mind, we decided to remove all the time series whose range is less than or equal to 10%. This reduced the number of time series from 802 to 281. The data set visualization after the application of the filtering operation can be found in Figure 4.5. From this, we can appreciate and identify patterns and trends in the time series data set.

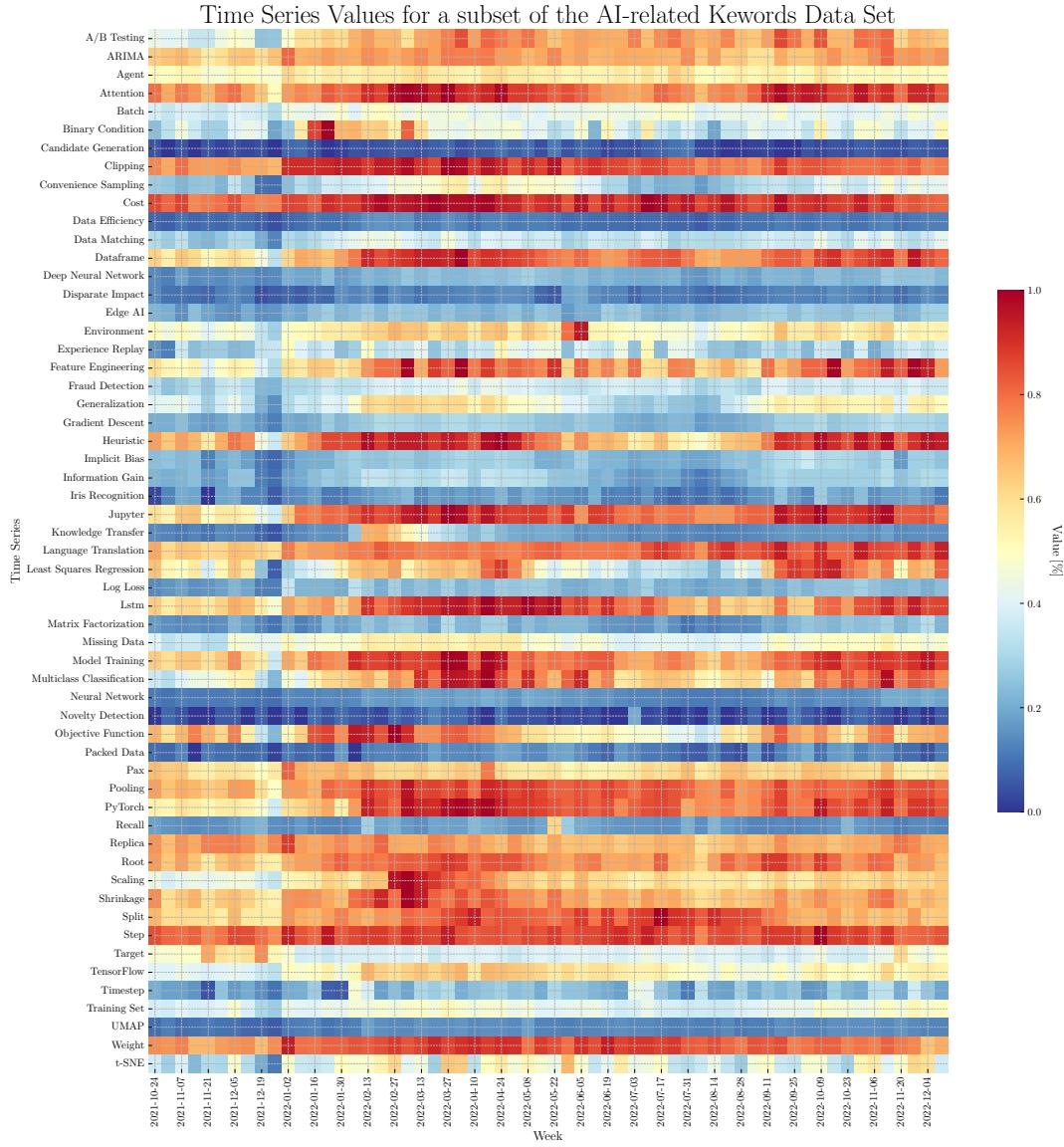


**Figure 4.5.:** Time Series plot for the 281 keywords obtained after filtering the original data set, included in Figure 4.4. The black portion of the lines indicates the training set, whereas the red portion of the lines defines the test set.

Nevertheless, given the sheer volume of time series under consideration, the line plot interpretation remains challenging. For that reason, we have prepared an alternative visualization on a subset of the data using tile plots, which can be found in Figure 4.6. The stark contrast between the adjacent squares representing the time series values allows us to conclude that this data set is rather noisy and that the forecasting task will be challenging. In particular, the observed noise is to be interpreted as *aleatoric uncertainty*, or uncertainty which is intrinsic in the data. As a result, such uncertainty must be properly accounted for through the use of prediction intervals, as the mean of the predictive distribution may not be informative enough and may only provide a very limited piece of information about the forecast.

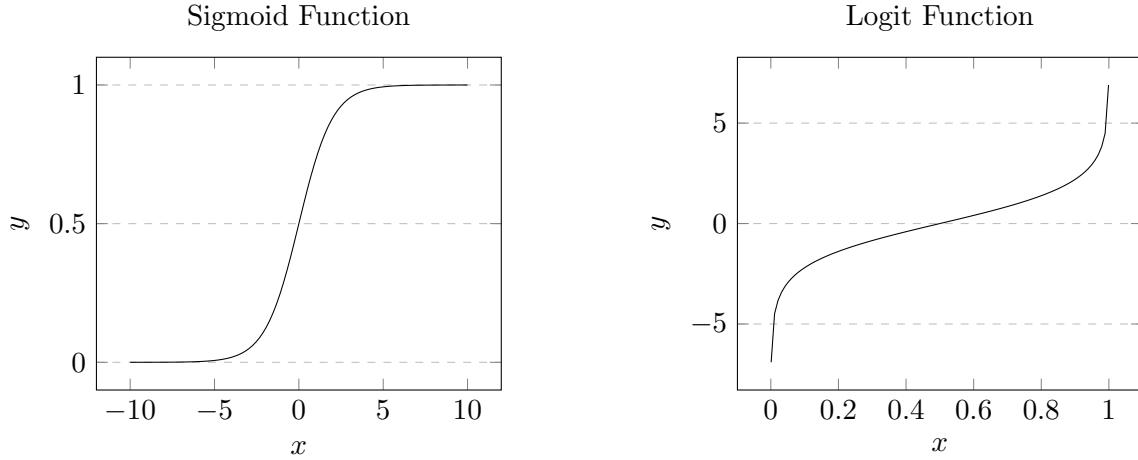
## 4.3. Scaling and Windowing

Many statistical models may struggle to provide predictions in a bounded interval. As a pre-training step, we first rescale the time series values to lie in the  $[0, 1]$  interval, we then clip this interval in the range  $[10^{-4}, 1 - 10^{-4}]$  to avoid undefined values before applying the logit transformation (right side in Figure 4.7). As the logit function maps from the  $(0, 1)$  interval to



**Figure 4.6.:** Visualization of a subset of the 281 keywords obtained after filtering the original data set. Hot colors represent high values whereas cold colors represent low values.

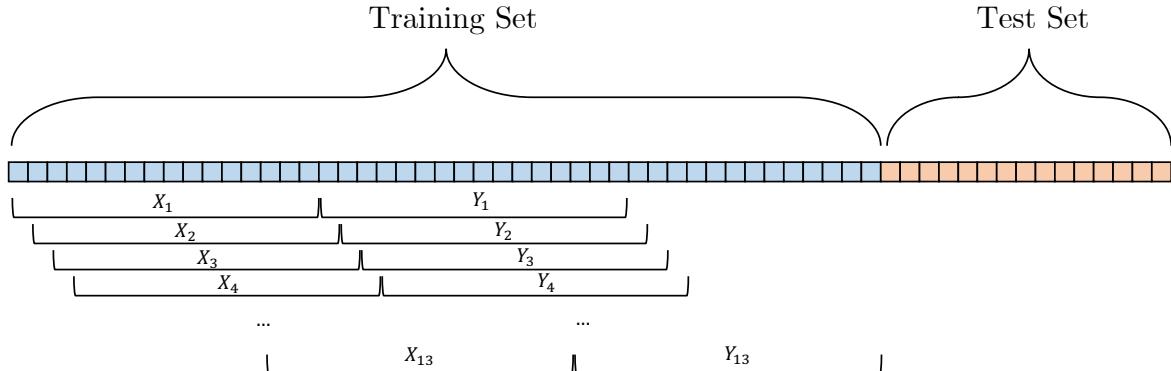
the real line,  $\text{logit} : (0, 1) \rightarrow \mathbb{R}$ , the issues concerning bounded predictions are solved, as long as we apply the inverse logit (more commonly known as sigmoid or logistic function, on the left in Figure 4.7) before evaluating the model's error. Since sigmoid and logit functions are strictly monotonic, the inverse transformation allows us to obtain some prediction intervals in the original space in a straightforward manner. However, it is crucial to note that while the sigmoid transformation is monotonic and preserves the relative ordering of the prediction intervals - if a model  $\mathcal{M}_A$  has wider prediction intervals than a second model  $\mathcal{M}_B$  in the transformed space, it will similarly have wider prediction intervals in the original space as well - it may not preserve the exact coverage probability or interval symmetry. Thus, while the comparison between prediction intervals for different models is allowed, particular care must be paid to the coverage probability in the original space. From this moment on, when we refer to 95% prediction intervals, please note we are referring to the application of the



**Figure 4.7.:** Plots of sigmoid  $\sigma(x) = \frac{1}{1 + e^{-x}}$  and logit( $x$ ) =  $\log\left(\frac{x}{1-x}\right)$  functions.

inverse transformation to the 95% prediction intervals in the transformed space.

Various strategies exist for calculating prediction intervals after the inverse transformation, with many of them relying on sampling or Delta method. However, as the underlying predictive distribution for ML/DL methods is not predetermined, estimates through sampling are unattainable. Nevertheless, using the inverse transformation offers a consistent comparison across all methods.



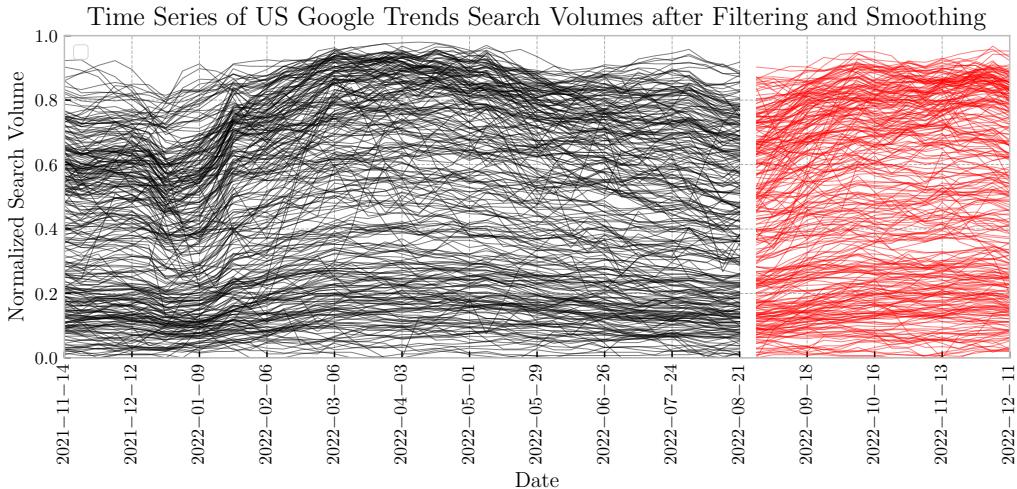
**Figure 4.8.:** Data set creation for machine learning and deep learning-based methods. Input and output window sizes are set to be equal to 16 weeks. The test set is not utilized during the training phase, but only for the final evaluation.

Finally, training of machine learning and deep learning models requires the prior chunking of the training data, as represented in Figure 4.8. The training data is split into overlapping windows comprising 16 weeks of inputs and 16 weeks of outputs. The reason for the specific choice of 16 weeks is related to the test window size, which has been chosen to be equal to 16 weeks as well in order to cover a full business quarter. By deciding an equal window size for inputs and outputs, we are effectively assuming that the input data encodes enough information to predict future values.

## 5. Results and Discussion

In this chapter, we will explore the results obtained by applying the methods described in Section 3.3 to the data set created in Chapter 4, with a further addition. Given the limited data we are considering, and the fact that the original data set is rather noisy, we want to make sure that the obtained results generalize to a smoother data set. In order to do this, we create a copy of the original (raw) data set and apply a 4-week-long moving average to the time series composing it. Such smoothed data set can be found in Figure 5.1 (cf. Figure 4.5). The reason for choosing a 4-week window can be linked to the way Google updates its core search engine algorithm. In particular, core updates are performed across a period of one month, and most years include only one or two core updates. Consequently, averaging over a 4-week period would automatically smooth out the sudden changes in the time series values which are the results of a change in the way Google counts the searches, rather than in a drop/increase of interest from the public.

The following two sections detail the results on the raw (Figure 4.5) and the smoothed (Figure 5.1) data sets.



**Figure 5.1.:** Time Series plot for the 281 keywords obtained after filtering the original data set and applying a 4-week moving average to the individual time series.

### 5.1. Raw Data Set Results

We begin by analyzing the test set results for the raw data set. Figure 5.2 contains information on the MAE incurred by the different methods across the 16-week forecast horizon (solid black line). The meaning of the models' names and the number of fitted models for each of the graphs are included in Table 5.1. The MAE alone should not be considered a reliable metric to select the best model, as it does not provide information on the spread of the magnitude the error may assume across the entirety of the data set. We instead

obtain the 0.025 and the 0.975 quantiles of the absolute errors - with respect to the mean prediction - and average these quantiles over 1000 bootstrapped data sets obtained from the original one (where in this case the bootstrapped data sets are obtained by sampling with replacement a number of time series equal to the one composing the original data set). These results are included in the graphs as dashed red lines. The same bootstrap approach is also used to calculate the confidence interval for the MAE across the forecast horizon, and, as expected, such confidence intervals (dashed black lines) are rather tight around the observed mean MAE (solid black line). By looking at the plots, it is plausible to conclude that that GP-WBIC and WLCM are the best-performing models in the entire set, as the upper bound for the error (upper dashed red line) is rather low across the 16-week horizon.

Establishing a baseline to compare the time series performance metrics is crucial: without a simple model to be used for comparison, it would be hard to state with a high degree of confidence that the proposed models do perform better than random guessing. In particular, if the time series is very noisy - to the point of being almost completely random - a simple model assuming as forecast the last observed value - also called a *naive* forecast - could prove rather powerful. The graph on the first row and first column in Figure 5.2 shows the results for the naive forecast.

Model	Description	Number of Models
Naive	Naive forecast	1 per series
ARIMA	Autoregressive integrated moving average	1 per series
Holt-Winters	Holt-Winters forecasting	1 per series
LGBM-Local	LightGBM model applied locally	1 per series/step/quantile
LSTM-Local	Long-Short Term Memory network applied locally	1 per series
CNN-Local	Convolutional Neural Network applied locally	1 per series
GP-BIC	Gaussian Process with BIC-based automatic kernel discovery	1 per series
<b>GP-WBIC</b>	Gaussian Process with WBIC-based automatic kernel discovery	1 per series
LGBM-Global	LightGBM model applied globally	1 per data set
LSTM-Global	Long-Short Term Memory network applied globally	1 per data set
CNN-Global	Convolutional Neural Network applied globally	1 per data set
<b>WLCM</b>	Weak Learner combined with a Clustering Model	1 per series/cluster/ $\lambda$
<b>WLRCM</b>	Weak Learner combined with a Residual Clustering Model	1 per series/cluster/ $\lambda$
<b>WLGM</b>	Weak Learner combined with a Global Model	1 per series/ $\lambda$

**Table 5.1.:** Summary of models, their descriptions, and the number of models used. Models in **bold** are original contributions of this research.

Figure 5.2 illustrates that most of the models exhibit a slight MAE trend or none at all. This implies that long-term forecasting is viable for this data set. However, a few models, including the naive model and the Holt-Winters model, behave differently. Notably, the naive model stands out, slightly outperforming the Holt-Winters model throughout the entire forecast horizon. A similar analysis can be conducted against the other trained models.

A table including the forecast MAE (see Definition 3.1.4) and AE975 (see Definition 3.1.6) is included in Table 5.2. For the purpose of this study, we have decided to average the MAE and the AE975 across the entire forecast horizon for each model and to obtain the metrics' ratio with respect to the naive forecast. In particular, AE975 is more crucial for the decision on which model is the best performing one, as it represents a reliable estimate of the maximum error to be observed on the data set.

Globally trained ML/DL models (LGBM-Global, LSTM-Global) with the exception of CNN-Global offer no benefit in comparison with the naive forecast, with all the metrics' ratios being higher than 1. The reason for their poor performance can be found in the inability of a single model to efficiently capture the idiosyncrasies (i.e. trend and cyclic patterns) of all time series simultaneously, focusing too much on the long-term patterns and

missing the short-term ones. CNN-Global, instead, does just the opposite.

Some of the ML/DL local models (LGBM-Local and LSTM-Local) slightly outperform the naive forecast in terms of both MAE and AE975. CNN-Local, on the other hand, is the worst performing among the analyzed models, which can be due to training set overfitting effects. Recall that the architecture optimization for the DL models had been performed in the context of the global models since attempting multiple architectures fittings on each time series would make the problem computationally intractable given the available resources.

While ARIMA offers slight benefits in comparison to the naive model, Holt-Winters perform poorly on the data set. This might be due to the amount of aleatoric uncertainty in the data, which forces the optimization routine behind the smoothing parameters' estimation in Holt-Winters to choose values leading to a reduced sensitivity to noise, thus making the model's "inertia" larger (slower reaction to trend changes in the time series).

GP-BIC and GP-WBIC perform well, with GP-WBIC (originally proposed in this thesis) improving over GP-BIC and achieving an MAE ratio of 0.775 (-22.5% with respect to the naive model) and an AE975 ratio of 0.785 (-21.5% with respect to the naive model). The magnitude of the improvement with respect to the GP method based on the BIC is not negligible either, with the MAE and AE975 ratios improving by respectively 20% and 48.5% (relative to the naive forecast). We believe that the improvement experienced by the GP with WBIC can be explained by the more effective way WBIC performs model selection, which considers the entire distribution of the parameters instead of a point metric. Furthermore, such results support the validity of the Monte Carlo approximation for the WBIC we provided in Equation 3.3.

The other novel models we introduced in this research (WLGM, WLCM, WLRCM) all perform better than the naive model on the data set. However, a more sensible comparison would be, in this case, using GP-WBIC as a baseline, since all these models use, for the weak learner, the differences between the original time series and the in-sample prediction from GP-WBIC weighted by a parameter  $\lambda$  (while there are alternatives available, we choose to construct all subsequent models upon the GP-WBIC, as it represents our primary contribution). The use of clustering in WLCM seems to have a beneficial effect on the AE975 metric, whose ratio decreases in comparison to GP-WBIC (-0.9%), although at the cost of a slight increase in the MAE ratio (+1.7%). WLGM is the worst-performing model among those proposed, which reinforces our belief in the existence of underlying cluster structures in the data.

## 5.2. Smoothed Data Set Results

The smooth data set is obtained by applying a rolling mean with a window of size 4 weeks to the raw data set. The resulting smoothed data set is shown in Figure 5.1, while the metrics' summary is provided in Table 5.3, and the plot of the MAE across the considered horizon is included in Figure 5.3.

In the context of the smoothed data set, we find that GP-WBIC, WLCM, and WLRCM provide the best performance. In particular, we observe that GP-WBIC is the overall best performer in terms of both MAE ratio (0.643) and AE975 ratio (0.597). On the other hand, WLCM is the second-best performer for the AE975 ratio (0.606) and WLRCM for the MAE ratio (0.697). Therefore, contrary to the noisy setting, a smoother data set does not demand a trade-off between MAE and AE975.

Once again, we observe that GP-WBIC dominates GP-BIC in terms of both MAE and AE975. In particular, while the decrease in MAE ratio is relatively small (0.195), the AE975

Model	MAE	MAE/Naive MAE	AE975	AE975/Naive AE975
Naive	0.080	1.000	0.380	1.000
ARIMA	0.072	0.902	0.372	0.979
Holt-Winters	0.093	1.170	0.507	1.335
LGBM-Local	0.068	0.856	0.347	0.914
LSTM-Local	0.071	0.896	0.355	0.935
CNN-Local	0.099	1.239	0.452	1.191
GP-BIC	0.082	1.029	0.482	1.270
GP-WBIC	<b>0.062</b>	<b>0.775</b>	0.298	0.785
LGBM-Global	0.080	1.008	0.389	1.024
LSTM-Global	0.083	1.048	0.399	1.050
CNN-Global	0.067	0.841	0.320	0.843
WLCM	<i>0.063</i>	<i>0.792</i>	<b>0.289</b>	<b>0.760</b>
WLRCM	0.066	0.834	0.319	0.840
WLGM	0.073	0.911	0.349	0.918

**Table 5.2.:** MAE and AE975 are provided for the used models fit on the raw data set. Relative performance is compared to the naive model (values lower than 1 signify that the model outperforms the naive forecast on the given metric). The numbers in **bold** indicate the best model according to the relative metric, whereas numbers in *italics* highlight the second best model for a given metric.

ratio goes from 1.315 for the GP-BIC to 0.597 for GP-WBIC, which translates into a maximum observed error reduction by more than 50% (from 0.482 to 0.218 in absolute terms).

ARIMA provides a marginal improvement in terms of MAE in comparison to the naive model, but its AE975 ratio is higher than 1, implying a higher chance of large errors than the naive model.

Additionally, the performance of the Holt-Winters model is rather poor (46.1% relative increase in the MAE with respect to the naive model). This could be due to the model's inability to capture non-linear trends owing to the employed additive formulation (refer to Section 3.3.2 for details on this point).

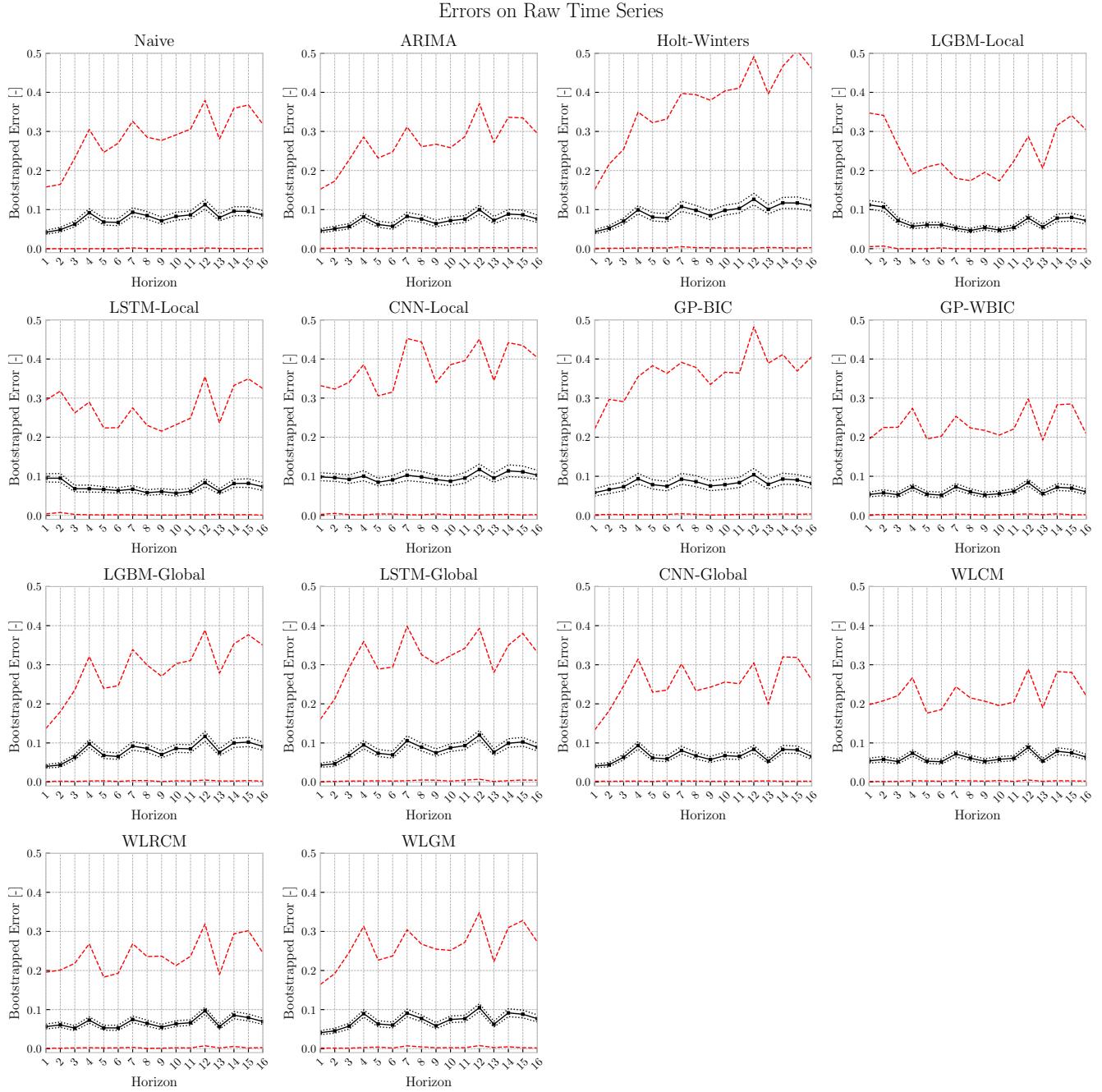
Some of the ML/DL models trained on the individual time series (LGBM-Local, LSTM-Local) perform better than the naive model while CNN-Local exhibits an increase in both MAE and AE975.

For the globally trained models (LGBM-Global, LSTM-Global, CNN-Global) the situation appears to be specular to the local case, with CNN-Global outperforming both LGBM-Global and LSTM-Global.

Finally, WLGM ranks very well in terms of both MAE and AE975. However, its performance does not come close to those of GP-WBIC, WLCM, and WLRCM, especially concerning AE975. This observation further supports the idea of underlying cluster structures in the data, which seem to be effectively captured by the DTW-HDBSCAN pairing and completely ignored by WLGM.

### 5.3. Noise Effect

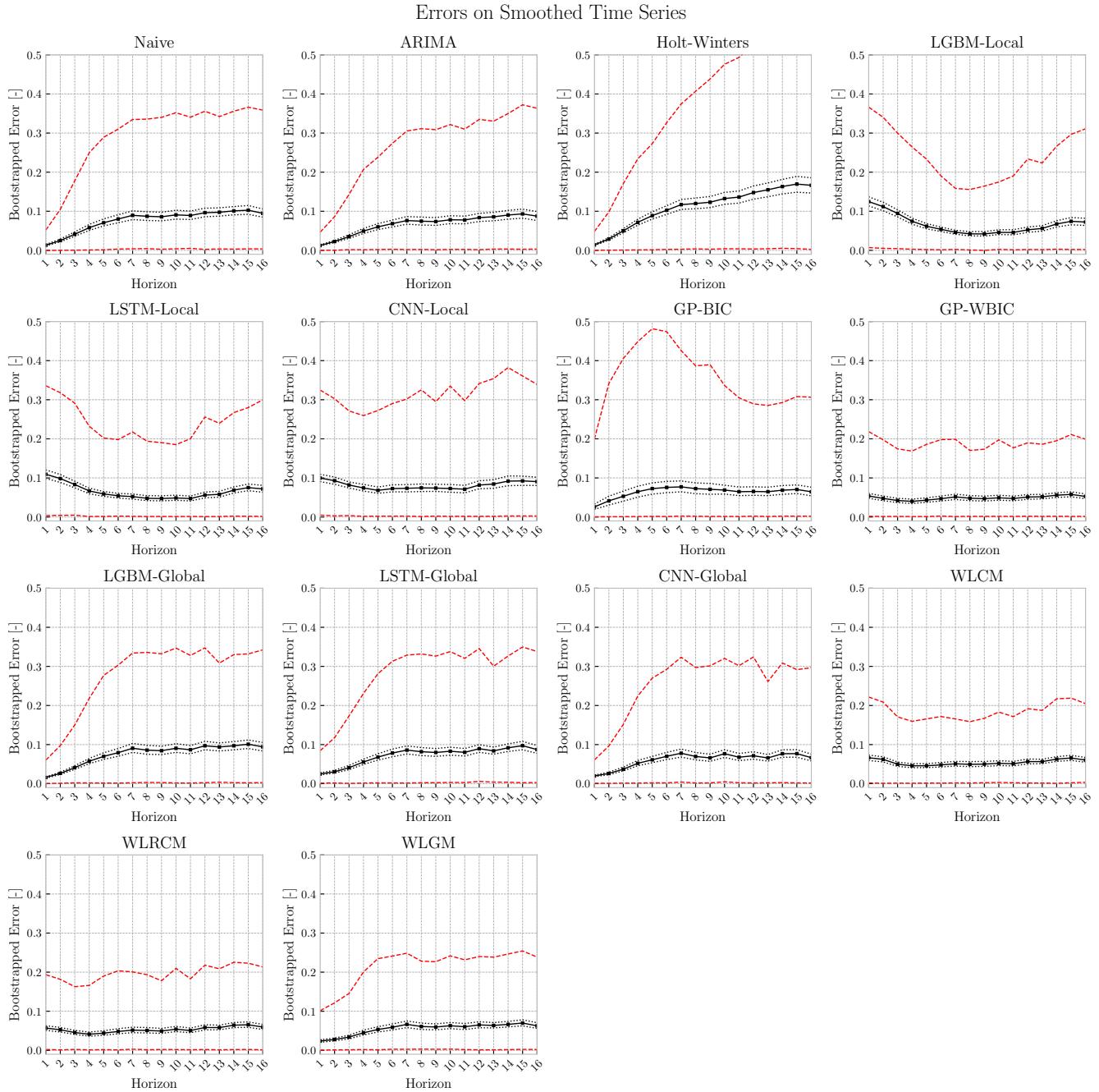
The application of a moving average of 4 weeks to the time series seems to be leading - with the exclusion of a few models - to a reduction in the MAE/AE975 metric. While one might expect complex models to perform poorly on noisy datasets, focusing more on the



**Figure 5.2.:** Plots of average errors for the raw time series over the 16-week test set. The solid black line indicates the mean absolute error, the dashed black lines indicate the confidence interval for the MAE estimate, and the dashed red lines indicate respectively the 0.025 and the 0.975 quantiles of the absolute errors for each value of the horizon.

noise than on the underlying signal, our findings do not consistently support this hypothesis. The Holt-Winters model, for example, often praised for its simplicity, exhibits an even worse performance than the more complex ML/DL models. We can therefore conclude that:

- noise can affect the performance of the forecasting models, but the magnitude of such



**Figure 5.3.:** Plots of average errors for the smoothed time series over the 16-week test set. The solid black line indicates the mean absolute error, the dashed black lines indicate the confidence interval for the MAE estimate, and the dashed red lines indicate respectively the 0.025 and the 0.975 quantiles of the absolute errors for each value of the horizon.

influence depends on both the data and the model and not on the data alone.

- a model robust to data noise should perform decently on a noise-free data set, but such a conclusion must be verified whenever possible.

Model	MAE	MAE/Naive MAE	AE975	AE975/Naive AE975
Naive	0.076	1.000	0.366	1.000
ARIMA	0.066	0.871	0.372	1.017
Holt-Winters	0.112	1.461	0.613	1.675
LGBM-Local	0.066	0.871	0.366	1.000
LSTM-Local	0.065	0.854	0.336	0.917
CNN-Local	0.081	1.060	0.382	1.044
GP-BIC	0.064	0.838	0.482	1.315
GP-WBIC	<b>0.049</b>	<b>0.643</b>	<b>0.218</b>	<b>0.597</b>
LGBM-Global	0.076	0.989	0.347	0.949
LSTM-Global	0.072	0.948	0.350	0.955
CNN-Global	0.061	0.802	0.324	0.885
WLCM	0.055	0.715	0.222	0.606
WLRCM	<i>0.053</i>	<i>0.697</i>	0.225	0.615
WLGM	0.055	0.722	0.254	0.695

**Table 5.3.:** MAE and AE975 are provided for the used models fit on the smoothed data set. Relative performance is compared to the naive model (values lower than 1 signify that the model outperforms the naive forecast on the given metric). The numbers in **bold** indicate the best model according to the relative metric, whereas numbers in *italics* highlight the second best model for a given metric.

## 5.4. Result comparison on a time series

Without any visual indication of what a time series forecast may look like, the comparison between the employed models becomes a mathematical exercise without clear interpretation. It is for this reason that we decided to select a time series and to show its 16-week forecast, as well as the prediction intervals that the models output. The results can be found in Figure 5.4.

We immediately notice that WLGM, WLCM, and WLRCM models only include a forecast but no prediction intervals. These models constitute new attempts introduced in this project, and prediction intervals cannot be easily produced for them in an analytical manner. We expect their development to be a topic for future research (possibly by working with individual posterior samples for the Gaussian Process, rather than with the forecast mean). They still provide robust forecasts *for the average* of the future realization.

The naive model forecasts the last observed value of the time series. We note that the width of the prediction interval expands quite rapidly, thus implying the high uncertainty this model would incur, and asymmetrically, owing to the application of the sigmoid transformation. The prediction intervals for the naive model are calculated by first differencing the time series (thus obtaining a differenced time series, which corresponds to the time series of the residuals) and then calculating the standard deviation of such residuals <sup>1</sup> which constitutes the width of half of the prediction interval for one time step ahead: we name this quantity  $\sigma_{\text{naive}}$ . The two-side 95% prediction interval can be then obtained by assuming normality of the residuals and finding:

$$q = \Phi^{-1}(0.975)$$

---

<sup>1</sup>The residuals are assumed to be independent and normally distributed. Without this assumption, we are not allowed to scale the prediction interval via the quantile of the normal distribution  $\Phi^{-1}(0.975)$  and we may not sum the variances of the distributions of the forecast to obtain the prediction interval at horizons larger than 1.

where  $\Phi^{-1}$  is the inverse of the standard normal cumulative distribution. Finally, the expansion of the prediction interval for larger horizons can be obtained by forecasting the error growth rate as increasing with the square root of the horizon  $h$ , thus leading to:

$$\sigma_{naive}^{95\%}(h) = \sigma_{naive}\Phi^{-1}(0.975)\sqrt{h}$$

which represents the size of the half-width of the 95% prediction interval for the naive model with increasing forecast horizon, prior to the application of the sigmoid transformation.

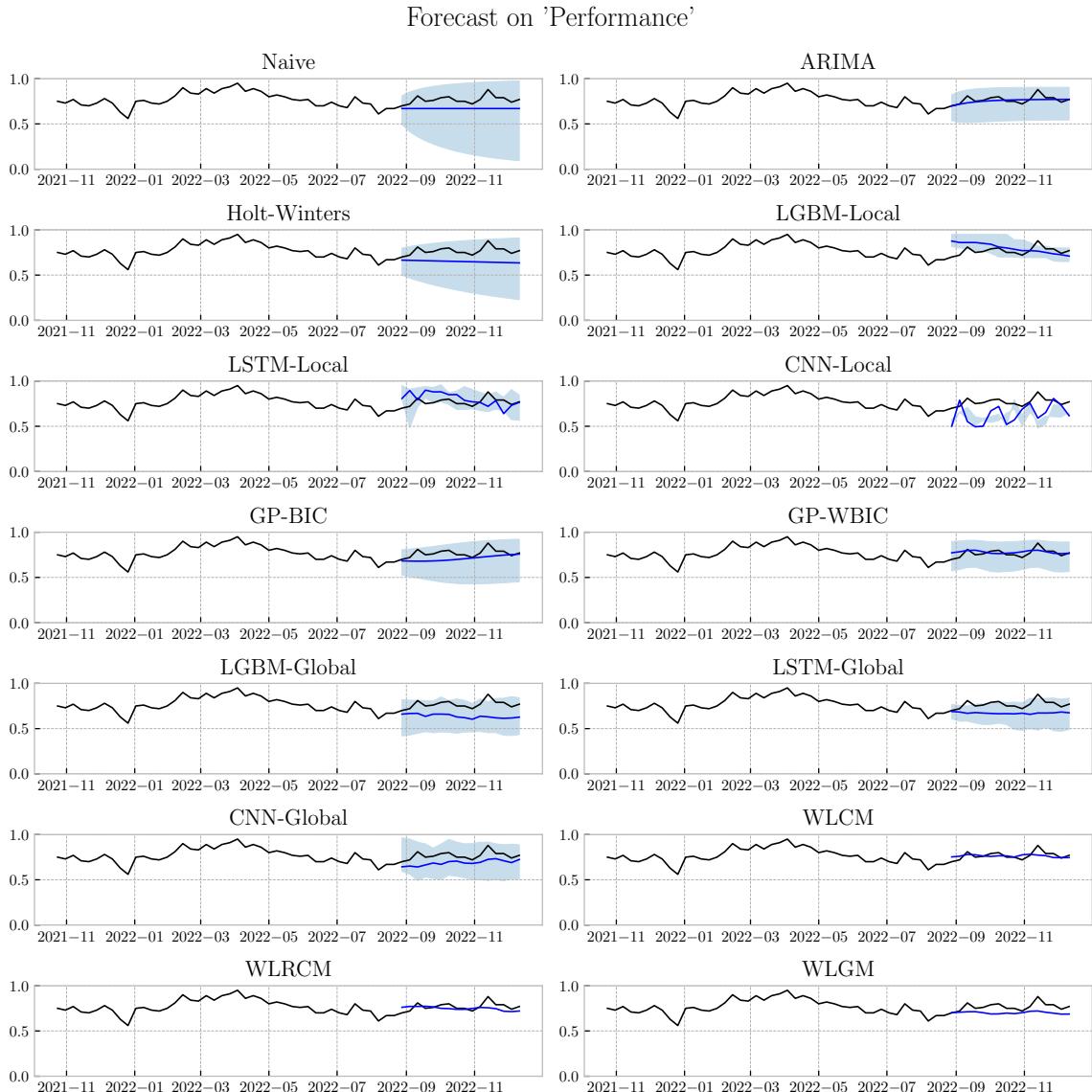
The Holt-Winters model yields a forecast similar to the naive model but has a marginally tighter prediction interval. In contrast, ARIMA outperforms in terms of both mean prediction and prediction interval tightness. The local ML/DL models, on average, struggle to capture the time series' correct trend. Notably, both the prediction intervals and the mean prediction of the CNN-Local model are off. For both local and global ML/DL models, the prediction intervals appear very noisy. This noise might result from the high variability of the data and the application of the pinball loss. While many analysts might opt to smooth such noisy prediction intervals for clarity, this approach can obscure the true nature of the model's forecast. We chose to retain this original representation.

Finally, GP-BIC and GP-WBIC provide smooth confidence intervals which remain rather stable as the horizon increases, although GP-WBIC seems to be slightly better at capturing, through its mean, the required 16-week forecast.

## 5.5. Results' Summary

Our results show the superiority of the proposed GP-WBIC in comparison to GP-BIC, found in literature, on both mean and maximum error. In addition, the introduced WLGCM, WLCM, and WLRCM provide good average forecasts (at the cost of a lack of prediction intervals, which we are positive may be correctly accounted for in future research). Furthermore, we confirmed that purely ML/DL models may not be particularly suitable for such data sets and that ARIMA and Holt-Winters only provide marginal (if any) benefits in comparison to the much simpler naive model.

One of the main points of concern in the application of time series forecasting is the prior knowledge in statistics and machine learning the marketing professionals may need in order to successfully apply the listed models in the context of time series forecasting. While most classical models may need prior specifications or setups - for example, the selection of a multiplicative or an additive model, the decision between using raw or differenced time series, the period to be set - machine learning and deep learning models may require, on the other hand, a trial-and-error procedure across an abundance of models to select the most appropriate one, something which may be both computationally expensive and cumbersome to implement. Using GP-WBIC provides high mean accuracy and prediction intervals, and it is completely automatic in the selection of the appropriate kernels for the time series fitting, therefore any marketing professional, with or without prior exposure to time series modeling, may confidently employ such model and even inspect its kernels, should they wish to explain the logic behind a given forecast. This paves the way for a more balanced distribution of the workload in the SEO field at any marketing company. While WLGM, WLCM, and WLRCM do not assume prior user knowledge of their underpinnings, they do not guarantee the same level of transparency GP-WBIC provides, since they leverage deep learning models which, in many circumstances, offer higher accuracy by sacrificing interpretability. In particular, if prediction intervals are of interest, we recommend that the user employ GP-WBIC.



**Figure 5.4.:** Plots of the ground truth (black line), forecast (blue line), and prediction intervals (blue region) for the utilized models on the keyword **Performance**. The raw time series is used. In all plots, the horizontal axes represent the date and the vertical axes represent the time series value.

## 5.6. Limitations

Although GP-WBIC, WLCM, and WLRCM have proven to provide highly accurate forecasts, the issue concerning the model's scalability remains open. As GP-WBIC needs to fit one Gaussian Process for some combinations of the base kernels, it incurs a computational cost of  $O(Ikn^3)$ , where  $k$  is the amount of kernels combinations attempted at each iteration (5 base kernels x 2 operations + 3 change point kernels),  $I$  is the total number of iterations, and  $n$  is the number of time steps.

$k$  may not be modified without heavily affecting the model's performance by restricting the search space, although this step can be easily parallelized in order to reduce the amount of computing time (in the context of this research, we parallelized the algorithm to train 30

time series simultaneously on 30 cores instead of parallelizing the training at the kernel level).

$I$  can vary from time series to time series, as the algorithm would automatically terminate once the WBIC stops decreasing (although a criterion with a maximum number of iterations may also be implemented). The model selection step is inherently serial (thus employing a machine with more than  $k = 13$  cores or threads would not speed up the training step, which is why we opted for the time series-parallel implementation instead of parallelizing the procedure at the kernel level).

$n$  is the parameter that most affects the overall computational cost. Unfortunately, as it is, the  $O(n^3)$  complexity lies in the statistical framework behind Gaussian Processes and not in their implementation, therefore not much can be done to reduce such cost. At the same time, we argue that a marketing professional is often interested in the latest trends, thus newly found keywords receiving a lot of attention may have a relatively short search history, thus limiting the computational cost induced by  $n$ .

Finally, note that WLGM, WLCM, and WLRCM leverage the knowledge acquired by different time series (in literature, *cross-training*) to achieve higher accuracy across the entire data set. It is obvious that these models can only be applied in contexts where the data sets contain more than a single time series.

## 6. Conclusion and Further Work

In this project, we have analyzed the performances of multiple common models found in the time series forecasting literature, focusing specifically on classical and ML/DL models. Chapter 2 provided some background and a literature review on time series analysis. We proceeded to explain the metrics used for evaluation in Chapter 3, detailing the models commonly found in textbooks and introducing some novel approaches in the context of Gaussian Processes for time series forecasting. As no public data set was available for keyword normalized search volumes, in Chapter 4, we resorted to Google's machine learning glossary and ChatGPT-4 to provide useful keywords to be queried on Google Trends, thus constructing a data set conducive for a fair model evaluation. Finally, in Chapter 5, we applied the methods described in Chapter 3 to the data set created in Chapter 4. This allowed us to evaluate the proposed novel methods (GP-WBIC, WLGM, WLCM, WLRCM) in contrast to the traditional ones. Notably, we also generated a second data set by smoothing the first via a moving average, determining the models' performances in both scenarios. This analysis unveiled the superiority of GP-WBIC, WLCM, and WLRCM compared to the other models under investigation, even though this higher accuracy came at a higher computational cost.

Furthermore, the GP-WBIC may be employed even by inexperienced users, or by users lacking a solid background in statistics and machine learning. This is due to its ability to perform automatic kernel discovery and to select the best kernel based on the WBIC measure. This feature proves particularly beneficial for SEO; all a marketing professional would need to predict the popularity of a keyword (and thus the resources to invest in a certain topic) would be a Google Trends search history. Coupled with a basic knowledge of how to parse a time series file in Python, the remainder of the action - model selection, and kernel discovery - would be automatically executed by the program we provide. This minimizes the amount of time needed to produce robust and reliable forecasts.

In conclusion, we are confident that our results are robust to variations in the data set, such as the inclusion of other time series from Google Trends. However, we also believe that more testing is required in order to gauge the benefits the proposed algorithms would bring in the more general realm of time series forecasting. Although significant advancements have been made to speed up Gaussian Process training over the past decade, newer, more efficient, and more robust libraries need to be introduced. In particular, the quality of a forecast from a GP is tied to the optimal maximum likelihood estimate for its parameters. A suboptimal estimate inevitably leads to lackluster forecasts, underscoring the need for a more intelligent parameter initialization. An avenue for research that we strongly recommend includes the use of Variational Gaussian Processes (VGPs) in place of regular GPs for time series forecasting. VGPs, especially Sparse VGPs, can decrease the time complexity of the model by leveraging some inducing points instead of the full time series, and allow the inclusion of non-Gaussian likelihoods which may be advantageous in contexts where the marginal distribution of the data is known or may be easily inferred. As no marginal log-likelihood can be calculated for VGPs, a challenging yet persistent issue would be translating the WBIC in the context of variational inference. Can the Evidence Lower Bound (ELBO) serve as a direct replacement for the marginal log-likelihood in the calculation of the WBIC? How can we streamline

model selection in a similar manner to what has been achieved for GPs? Will the use of the variational approximation heavily affect accuracy? If so, is the reduced accuracy a worthwhile trade-off for significant computational time savings?

In wrapping up this thesis, it is evident that, while we have made progress in the field of time series forecasting, a vast landscape of opportunities - which we hope will beckon future researchers - awaits to be discovered and explored. As further advancement is achieved, we believe that the balance between computational efficiency and forecasting accuracy will forever remain of paramount importance.

# Bibliography

- Marta Bańbura, Domenico Giannone, and Lucrezia Reichlin. Large bayesian vector auto-regressions. *Journal of applied Econometrics*, 25(1):71–92, 2010.
- The World Bank. Individuals using the internet (% of population), 2023. URL <https://data.worldbank.org/indicator/IT.NET.USER.ZS?end=2021&start=2000>. Accessed on: August 23.
- Gianluca Bontempi, Souhaib Ben Taieb, and Yann-Aël Le Borgne. *Machine Learning Strategies for Time Series Forecasting*, pages 62–77. Springer Berlin Heidelberg, Berlin, Heidelberg, 2013.
- George E. P. Box, Gwilym M. Jenkins, Gregory C. Reinsel, Greta M. Ljung, and George E. P. Box. *Time Series Analysis : Forecasting and Control*. Wiley, Hoboken, New Jersey, 2016.
- George E.P. Box, Gwilym M Jenkins, and G Reinsel. Time series analysis: Forecasting and control. *Time Series Analysis: Forecasting and Control*, 1970.
- Leo Breiman. Random forests. *Machine learning*, 45:5–32, 2001.
- Peter J. Brockwell and Richard A. Davis. *Time Series Theory and Methods*. Springer, New York, 2 edition, 2006.
- Ricardo JGB Campello, Davoud Moulavi, Arthur Zimek, and Jörg Sander. Hierarchical density estimates for data clustering, visualization, and outlier detection. *ACM Transactions on Knowledge Discovery from Data (TKDD)*, 10(1):1–51, 2015.
- Manuel Castells. *The Rise of the Network Society, Volume 1*. Wiley Online Library, 2009.
- Christopher Chatfield. *The Analysis of Time Series : An Introduction*. Chapman and Hall/CRC, Boca Raton, FL ;, 6 edition, 2004.
- Kyunghyun Cho, Bart Van Merriënboer, Caglar Gulcehre, Dzmitry Bahdanau, Fethi Bougares, Holger Schwenk, and Yoshua Bengio. Learning phrase representations using rnn encoder-decoder for statistical machine translation. *arXiv preprint arXiv:1406.1078*, 2014.
- Youngseog Chung, Willie Neiswanger, Ian Char, and Jeff Schneider. Beyond pinball loss: Quantile methods for calibrated uncertainty quantification. *Advances in Neural Information Processing Systems*, 34:10971–10984, 2021.
- V. N. Coelho, I. M. Coelho, B. N. Coelho, M. J. F. Souza, F. G. Guimarães, E. J. d. S. Luz, A. C. Barbosa, M. N. Coelho, G. G. Netto, R. C. Costa, A. A. Pinto, A. d. P. Figueiredo, M. E. V. Elias, D. C. O. G. Filho, and T. A. Oliveira. Eeg time series learning and classification using a hybrid forecasting model calibrated with gvns. *Electronic Notes in Discrete Mathematics*, 58:79–86, 2017. ID: 273151.

- Colah. Understanding lstm networks, 2015. URL <https://colah.github.io/posts/2015-08-Understanding-LSTMs/>. [Online; accessed date].
- McKinsey & Company. The economic potential of generative ai: the next productivity frontier, 2023. URL <https://www.mckinsey.com/capabilities/mckinsey-digital/our-insights/the-economic-potential-of-generative-ai-the-next-productivity-frontier>. Accessed on: August 23.
- Corinna Cortes and Vladimir Vapnik. Support-vector networks. *Machine learning*, 20:273–297, 1995.
- Alysha M De Livera, Rob J Hyndman, and Ralph D Snyder. Forecasting time series with complex seasonal patterns using exponential smoothing. *Journal of the American statistical association*, 106(496):1513–1527, 2011.
- D. Dickey and Wayne Fuller. Distribution of the estimators for autoregressive time series with a unit root. *JASA. Journal of the American Statistical Association*, 74, 06 1979. doi: 10.2307/2286348.
- David Duvenaud. *Automatic Model Construction with Gaussian Processes*. PhD thesis, University of Cambridge, Pembroke College, 2014.
- David Duvenaud, James Lloyd, Roger Grosse, Joshua Tenenbaum, and Zoubin Ghahramani. Structure discovery in nonparametric regression through compositional kernel search. *Proceedings of the 30th International Conference on Machine Learning*, 02 2013.
- Jonas Gehring, Michael Auli, David Grangier, Denis Yarats, and Yann N Dauphin. Convolutional sequence to sequence learning. In *International conference on machine learning*, pages 1243–1252. PMLR, 2017.
- G.R. Grimmett and D.R. Stirzaker. *Probability and Random Processes*. Oxford University Press, 2001.
- J. Han, M. Kamber, and J. Pei. *Data Mining: Concepts and Techniques*. The Morgan Kaufmann Series in Data Management Systems. Elsevier Science, 2011. ISBN 9780123814807. URL <https://books.google.co.jp/books?id=pQws07tdpjoC>.
- Nick Heard. *An Introduction to Bayesian Inference, Methods and Computation*. Springer, 2021.
- Sepp Hochreiter and Jürgen Schmidhuber. Long short-term memory. *Neural computation*, 9(8):1735–1780, 1997.
- Charles C Holt. Forecasting trends and seasonals by exponentially weighted moving averages. *ONR Memorandum*, 52(52):5–10, 1957.
- Rob J. Hyndman and George Athanasopoulos. *Forecasting : Principles and Practice*. OTexts, Melbourne, 2021.
- Alan Jovic and Franjo Jovic. Classification of cardiac arrhythmias based on alphabet entropy of heart rate variability time series. *Biomedical Signal Processing and Control*, 31:217–230, 2017. ID: 273545.

- Guolin Ke, Qi Meng, Thomas Finley, Taifeng Wang, Wei Chen, Weidong Ma, Qiwei Ye, and Tie-Yan Liu. Lightgbm: A highly efficient gradient boosting decision tree. *Advances in neural information processing systems*, 30, 2017.
- Diederik P. Kingma and Max Welling. Auto-encoding variational bayes. *CoRR*, abs/1312.6114, 2013.
- Roger Koenker and Gilbert Bassett Jr. Regression quantiles. *Econometrica: journal of the Econometric Society*, pages 33–50, 1978.
- Gabriel Kronberger and Michael Kommenda. Evolution of covariance functions for gaussian process regression using genetic programming. In *Computer Aided Systems Theory-EUROCAST 2013: 14th International Conference, Las Palmas de Gran Canaria, Spain, February 10-15, 2013, Revised Selected Papers, Part I* 14, pages 308–315. Springer, 2013.
- Pierre Simon Laplace. *Théorie Analytique Des Probabilités*. Courcier, 1814.
- Neil D. Lawrence. Gaussian process latent variable models for visualisation of high dimensional data. In *Conference on Neural Information Processing Systems*, 2003.
- Yann LeCun, Bernhard Boser, John S Denker, Donnie Henderson, Richard E Howard, Wayne Hubbard, and Lawrence D Jackel. Backpropagation applied to handwritten zip code recognition. *Neural computation*, 1(4):541–551, 1989.
- Yann LeCun, Léon Bottou, Yoshua Bengio, and Patrick Haffner. Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, 86(11):2278–2324, 1998.
- Bryan Lim and Stefan Zohren. Time-series forecasting with deep learning: a survey. *Philosophical transactions of the Royal Society of London.Series A: Mathematical, physical, and engineering sciences; Philos Trans A Math Phys Eng Sci*, 379(2194):20200209, 2021.
- James Lloyd, David Duvenaud, Roger Grosse, Joshua Tenenbaum, and Zoubin Ghahramani. Automatic construction and natural-language description of nonparametric regression models. *Proceedings of the National Conference on Artificial Intelligence*, 2, 02 2014. doi: 10.1609/aaai.v28i1.8904.
- Spyros Makridakis, Evangelos Spiliotis, and Vassilios Assimakopoulos. Statistical and machine learning forecasting methods: Concerns and ways forward. *PloS one; PLoS One*, 13 (3):e0194889, 2018.
- Leland McInnes, John Healy, and Steve Astels. Hdbscan: Hierarchical density based clustering. *J. Open Source Softw.*, 2(11):205, 2017.
- Radford M Neal et al. Mcmc using hamiltonian dynamics. *Handbook of markov chain monte carlo*, 2(11):2, 2011.
- John A Nelder and Roger Mead. A simplex method for function minimization. *The computer journal*, 7(4):308–313, 1965.
- Andy Pole, Mike West, and Jeff Harrison. *Applied Bayesian Forecasting and Time Series Analysis*. Chapman and Hall/CRC, 2018.
- PwC. Sizing the prize, 2023. URL <https://www.pwc.com/gx/en/issues/data-and-analytics/publications/artificial-intelligence-study.html>. Accessed on: August 23.

- Carl E. Rasmussen and Christopher K. I. Williams. *Gaussian Processes for Machine Learning*. MIT Press, Cambridge, 2005. ISBN 026218253X. Available also in a print ed.
- S. Roberts, M. Osborne, M. Ebdon, S. Reece, N. Gibson, and S. Aigrain. Gaussian processes for time-series modelling. *Philosophical transactions of the Royal Society of London.Series A: Mathematical, physical, and engineering sciences; Proc.R.Soc.A*, 371(1984):20110550, 2013.
- David E Rumelhart, Geoffrey E Hinton, and Ronald J Williams. Learning representations by back-propagating errors. *nature*, 323(6088):533–536, 1986.
- H. Sakoe and S. Chiba. Dynamic programming algorithm optimization for spoken word recognition. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, 26(1):43–49, 1978. doi: 10.1109/TASSP.1978.1163055.
- Gideon Schwarz. Estimating the dimension of a model. *The annals of statistics*, pages 461–464, 1978.
- Steven L Scott and Hal R Varian. Predicting the present with bayesian structural time series. *International Journal of Mathematical Modelling and Numerical Optimisation*, 5 (1-2):4–23, 2014.
- Christopher A Sims. Macroeconomics and reality. *Econometrica: journal of the Econometric Society*, pages 1–48, 1980.
- Statista. Market share of leading desktop search engines worldwide from january 2015 to march 2023, 2023. URL <https://www.statista.com/statistics/216573/worldwide-market-share-of-search-engines/>. Accessed on: August 23.
- Sean J. Taylor and Benjamin Letham. Forecasting at scale. *The American Statistician*, 72 (1):37–45, 2018.
- Sumio Watanabe. A widely applicable bayesian information criterion. *The Journal of Machine Learning Research*, 14(1):867–897, 2013.
- Mike West and Jeff Harrison. *Bayesian Forecasting and Dynamic Models*. Springer Science & Business Media, 2006.
- Peter R Winters. Forecasting sales by exponentially weighted moving averages. *Management science*, 6(3):324–342, 1960.
- Daniela Witten, Trevor Hastie, Rovert Tibshirani, and Gareth James. *An Introduction to Statistical Learning with Applications in R, Second Edition*. Springer Publication, 2021.

# A. Keywords Lists

## A.1. Google Machine Learning Glossary

This section contains the keyword list obtained from the Google Machine Learning Glossary.

A/B Testing	Convex Function	Exploding Gradient Problem
Accelerator Chip	Convex Optimization	F1
Accuracy	Convex Set	Fairness Constraint
Action	Convolution	Fairness Metric
Activation Function	Convolutional Filter	False Negative (Fn)
Active Learning	Convolutional Layer	False Negative Rate
Adagrad	Convolutional Neural Network	False Positive (Fp)
Agent	Convolutional Operation	False Positive Rate (Fpr)
Agglomerative Clustering	Cost	Feature
Anomaly Detection	Co-Training	Feature Cross
Ar	Counterfactual Fairness	Feature Engineering
Area Under The Pr Curve	Coverage Bias	Feature Extraction
Area Under The Roc Curve	Crash Blossom	Feature Importances
Artificial General Intelligence	Critic	Feature Set
Artificial Intelligence	Cross-Entropy	Feature Spec
Attention	Cross-Validation	Feature Vector
Attribute	Data Analysis	Federated Learning
Attribute Sampling	Data Augmentation	Feedback Loop
Auc (Area Under The Roc Curve)	Dataframe	Feedforward Neural Network (Ffn)
Augmented Reality	Data Parallelism	Few-Shot Learning
Automation Bias	Data Set Or Dataset	Fiddle
Automl	Dataset Api (Tf.Data)	Fine Tuning
Auxiliary Loss	Decision Boundary	Flax
Average Precision	Decision Forest	Flaxformer
Axis-Aligned Condition	Decision Threshold	Forget Gate
Backpropagation	Decision Tree	Full Softmax
Bagging	Deep Model	Fully Connected Layer
Bag Of Words	Decoder	Gan
Baseline	Deep Neural Network	Generalization
Batch	Deep Q-Network (Dqn)	Generalization Curve
Batch Normalization	Demographic Parity	Generalized Linear Model
Batch Size	Denoising	Generative Adversarial Network (Gan)
Bayesian Neural Network	Dense Feature	Generative Model
Bayesian Optimization	Dense Layer	Generator
Bellman Equation	Depth	Gpt (Generative Pre-Trained Transformer)
Bert (Bidirectional Encoder Representations From Transformers)	Depthwise Separable Convolutional Neural Network (Sepcnn)	Gini Impurity
Bias (Ethics/Fairness)	Derived Label	Gradient
Bias (Math) Or Bias Term	Device	Gradient Boosting
Bigram	Dimension Reduction	Gradient Boosted (Decision) Trees (Gbt)
Bidirectional	Dimensions	Gradient Clipping
Bidirectional Language Model	Discrete Feature	Gradient Descent
Binary Classification	Discriminative Model	Graph
Binary Condition	Discriminator	Graph Execution
Binning	Disparate Impact	Greedy Policy
Bleu (Bilingual Evaluation Understudy)	Disparate Treatment	Ground Truth
Boosting	Divisive Clustering	Group Attribution Bias
Bounding Box	Downsampling	Hallucination
Broadcasting	Dqn	Hashing
Bucketing	Dropout Regularization	Heuristic
Calibration Layer	Dynamic	Hidden Layer
Candidate Generation	Dynamic Model	Hierarchical Clustering
Candidate Sampling	Eager Execution	Hinge Loss
Categorical Data	Early Stopping	Holdout Data
Causal Language Model	Earth Mover'S Distance (Emd)	Hyperparameter
Centroid	Edit Distance	Hyperplane
Centroid-Based Clustering	Einsum Notation	I.I.D.
Checkpoint	Embedding Layer	Image Recognition
Class	Embedding Space	Imbalanced Dataset
Classification Model	Embedding Vector	Implicit Bias
Classification Threshold	Empirical Risk Minimization (Erm)	Imputation
Class-Imbalanced Dataset	Encoder	Incompatibility Of Fairness Metrics
Clipping	Ensemble	Independently And Identically Distributed (I.I.D.)
Cloud Tpu	Entropy	Individual Fairness
Clustering	Environment	Inference
Co-Adaptation	Episode	Inference Path
Collaborative Filtering	Epoch	Information Gain
Condition	Epsilon Greedy Policy	In-Group Bias
Configuration	Equality Of Opportunity	Input Generator
Confirmation Bias	Equalized Odds	Input Layer
Confusion Matrix	Estimator	In-Set Condition
Continuous Feature	Example	Instance
Convenience Sampling	Experience Replay	Interpretability
Convergence	Experimenter'S Bias	Inter-Rater Agreement

Intersection Over Union (IoU)	Novelty Detection	Root Mean Squared Error (Rmse)
Iou	Numerical Data	Rotational Invariance
Item Matrix	Numpy	R-Squared
Items	Objective	Sampling Bias
Iteration	Objective Function	Sampling With Replacement
Jax	Oblique Condition	Savedmodel
Keras	Offline	Saver
Keypoints	Offline Inference	Scalar
Kernel Support Vector Machines (Ksvms)	One-Hot Encoding	Scikit-Learn
K-Fold Cross Validation	One-Shot Learning	Scoring
K-Means	One-Vs.-All	Selection Bias
K-Median	Online	Self-Attention (Also Called Self-Attention Layer)
L0 Regularization	Online Inference	Self-Supervised Learning
L1 Loss	Operation (Op)	Self-Training
L1 Regularization	Optax	Semi-Supervised Learning
L2 Loss	Out-Of-Bag Evaluation (Oob Evaluation)	Sensitive Attribute
L2 Regularization	Optimizer	Sentiment Analysis
Label	Out-Group Homogeneity Bias	Sequence Model
Labeled Example	Outlier Detection	Sequence-To-Sequence Task
Label Leakage	Outliers	Serving
Lambda (Language Model For Dialogue Applications)	Output Layer	Shape (Tensor)
Lambda	Overfitting	Shrinkage
Landmarks	Oversampling	Sigmoid Function
Language Model	Packed Data	Similarity Measure
Large Language Model	Pandas	Single Program / Multiple Data (Spmd)
Layer	Parameter	Size Invariance
Layers Api (Tf.Layers)	Parameter Server (Ps)	Sketching
Leaf	Parameter Update	Softmax
Learning Rate	Partial Derivative	Sparse Feature
Least Squares Regression	Participation Bias	Sparse Representation
Linear Model	Partitioning Strategy	Sparse Vector
Linear	Pax	Sparsity
Linear Regression	Perceptron	Spatial Pooling
Logistic Regression	Performance	Split
Logits	Permutation Variable Importances	Splitter
Log Loss	Perplexity	Spmd
Log-Odds	Pipeline	Squared Hinge Loss
Long Short-Term Memory (Lstm)	Pipelining	Squared Loss
Loss	Pjit	Staged Training
Loss Aggregator	Pmap	State
Loss Curve	Policy	State-Action Value Function
Loss Function	Pooling	Static
Loss Surface	Positional Encoding	Static Inference
Lstm	Positive Class	Stationarity
Machine Learning	Post-Processing	Step
Majority Class	Pr Auc (Area Under The Pr Curve)	Step Size
Markov Decision Process (Mdp)	Praxis	Stochastic Gradient Descent (Sgd)
Markov Property	Precision	Stride
Masked Language Model	Precision-Recall Curve	Structural Risk Minimization (Srm)
Matplotlib	Prediction	Subsampling
Matrix Factorization	Prediction Bias	Summary
Mean Absolute Error (Mae)	Predictive Parity	Supervised Machine Learning
Mean Squared Error (Mse)	Predictive Rate Parity	Synthetic Feature
Metric	Preprocessing	T5
Meta-Learning	Pre-Trained Model	T5X
Metrics Api (Tf.Metrics)	Prior Belief	Tabular Q-Learning
Mini-Batch	Probabilistic Regression Model	Target
Mini-Batch Stochastic Gradient Descent	Proxy (Sensitive Attributes)	Target Network
Minimax Loss	Proxy Labels	Task
Minority Class	Pure Function	Temporal Data
MI	Q-Function	Tensor
Mnist	Q-Learning	Tensorboard
Modality	Quantile	Tensorflow
Model	Quantile Bucketing	Tensorflow Playground
Model Capacity	Quantization	Tensorflow Serving
Model Parallelism	Queue	Tensor Processing Unit (Tpu)
Model Training	Random Forest	Tensor Rank
Momentum	Random Policy	Tensor Shape
Multi-Class Classification	Ranking	Tensor Size
Multi-Class Logistic Regression	Rank (Ordinality)	Tensorstore
Multi-Head Self-Attention	Rank (Tensor)	Termination Condition
Multimodal Model	Rater	Test
Multinomial Classification	Recall	Test Loss
Multinomial Regression	Recommendation System	Test Set
Multitask	Rectified Linear Unit (Relu)	Text Span
Nan Trap	Recurrent Neural Network	Tf.Example
Natural Language Understanding	Regression Model	Tf.Keras
Negative Class	Regularization	Threshold (For Decision Trees)
Negative Sampling	Regularization Rate	Time Series Analysis
Neural Architecture Search (Nas)	Reinforcement Learning (Rl)	Timestep
Neural Network	Relu	Token
Neuron	Replay Buffer	Tower
N-Gram	Reporting Bias	Tpu
Nlu	Representation	Tpu Chip
Node (Neural Network)	Re-Ranking	Tpu Device
Node (Tensorflow Graph)	Retrieval-Augmented Generation	Tpu Master
Node (Decision Tree)	Return	Tpu Node
Noise	Reward	Tpu Pod
Non-Binary Condition	Ridge Regularization	Tpu Resource
Nonlinear	Rnn	Tpu Slice
Non-Response Bias	Roc (Receiver Operating Characteristic) Curve	Tpu Type
Nonstationarity	Root	Tpu Worker
Normalization	Root Directory	

Training	Undersampling	Variational Autoencoder (Vae)
Training Loss	Unidirectional	Wasserstein Loss
Training-Serving Skew	Unidirectional Language Model	Weight
Training Set	Unlabeled Example	Weighted Alternating Least Squares (Wals)
Trajectory	Unsupervised Machine Learning	Weighted Sum
Transfer Learning	Uplift Modeling	Wide Model
Transformer	Upweighting	Width
Translational Invariance	User Matrix	Wisdom Of The Crowd
Trigram	Validation	Word Embedding
True Negative (Tn)	Validation Loss	Xla (Accelerated Linear Algebra)
True Positive (Tp)	Validation Set	Zero-Shot Learning
True Positive Rate (Tpr)	Value Imputation	Z-Score Normalization
Unawareness (To A Sensitive Attribute)	Vanishing Gradient Problem	
Underfitting	Variable Importances	

## A.2. ChatGPT-4

This section contains the keyword list obtained from ChatGPT-4.

AI Ethics	Data Fusion	Information Extraction
AI Governance	Data Governance	Information Retrieval
AI for Social Good	Data Imputation	Interpretable Machine Learning
AI in Agriculture	Data Integration	Intrusion Detection Systems
AI in Cybersecurity	Data Labeling	IoT and AI
AI in Drug Discovery	Data Lakes	Iris Recognition
AI in Earth Sciences	Data Lineage	Jupyter
AI in Education	Data Matching	K-Means
AI in Finance	Data Privacy	K-Nearest Neighbors
AI in Gaming	Data Quality	Keras
AI in Genomics	Data Synthesis	Knowledge Distillation
AI in Healthcare	Data Transformation	Knowledge Graphs
AI in Manufacturing	Data Visualization	Knowledge Transfer
AI in Retail	Data Warehousing	L1 Regularization
AI in Robotics	Decision Trees	L2 Regularization
ARIMA	Deep Learning	Language Translation
AUC-ROC	Deep Q-Networks	Large Language Models
Accuracy	Deep Reinforcement Learning	Latent Dirichlet Allocation
Action Recognition	DeepMind	Learning Analytics
Active Learning	Differential Privacy	LightGBM
Actor-Critic Methods	Dimensionality Reduction	Linear Regression
Adversarial Attacks	Document Analysis	Link Prediction
Algorithmic Bias	Domain Adaptation	Logistic Regression
Anomaly Detection	ETL	Long Short-Term Memory Networks
Anomaly Detection in Time Series	Edge AI	Machine Learning
Artificial Intelligence	Educational Data Mining	Malware Detection
Attention Maps	Efficient Neural Networks	Market Basket Analysis
Attention Mechanism	Emotion Recognition	Matplotlib
Audio Classification	Ensemble Learning	Matrix Factorization
Audio Signal Processing	Evaluation Metrics	Medical Image Analysis
Audio Source Separation	Evolutionary Algorithms	Mental Health and AI
Augmented Reality and AI	Explainable AI	Meta Learning
Autoencoders	Explainable Deep Learning	Missing Data
Automated Machine Learning	F1 Score	Mixed Reality and AI
BERT	Face Recognition	Mobile AI
Bayesian Optimization	Fairness in Machine Learning	Model Compression
Bias Correction	Feature Engineering	Model Pruning
Bias in NLP	Feature Extraction	Model Robustness
Bias-Variance Tradeoff	Feature Selection	MongoDB
Big Data	Federated Learning	Monte Carlo Methods
Binary Classification	Few-Shot Learning	Motion Capture
Biometrics and AI	Fingerprint Recognition	Motion Planning
Blockchain and AI	Forecasting	Multi-agent Systems
Bootstrap Aggregating	Fraud Detection	Multi-modal Learning
Cassandra	GANs for Art Generation	Multi-task Learning
CatBoost	GANs for Data Augmentation	Multiclass Classification
Causal Inference	Game Playing AI	Multilingual Models
Chatbots	Game Theory and AI	Music Generation
Churn Prediction	Generative Adversarial Networks	NLP for Low-resource Languages
Climate Modeling	Generative Pre-trained Transformer	Named Entity Recognition
Clustering	Genetic Algorithms	Natural Language Processing
Collaborative Filtering	Geospatial Analysis	Neural Architecture Search
Computer Vision	GloVe	Neural Networks
Confusion Matrix	Gradient Boosting	Neural Style Transfer
Content-Based Filtering	Graph Mining	Neuro-Linguistic Programming
Contrastive Learning	Graph Neural Networks	Neuromorphic Computing
Conversational AI	Grid Search	NoSQL
Convolutional Neural Networks	Hadoop	Numpy
Cross-Validation	Hardware Accelerators	Object Detection
Cross-modal Learning	Healthcare Predictive Analytics	One-Shot Learning
Crowdsourcing	Hierarchical Clustering	OpenAI
Customer Analytics	High-Performance Machine Learning	OpenAI GPT Models
Customer Sentiment Analysis	Human Activity Recognition	Optical Character Recognition
CycleGAN	Human-in-the-loop Machine Learning	Out-of-Distribution Detection
DBSCAN	Hyperparameter Tuning	Overfitting
Data Augmentation	Image Captioning	Pandas
Data Bias	Image Classification	Parallel and Distributed Machine Learning
Data Cleaning	Image Segmentation	Particle Swarm Optimization
Data Compression	Image Super-Resolution	Personalized Learning
Data Efficiency	Imbalanced Data	Policy Gradients

Pose Estimation	SQL	Synthetic Minority Over-sampling Technique
Precision	Saliency Maps	T-SNE
Predictive Maintenance	Satellite Imagery Analysis	TensorFlow
Principal Component Analysis	Scalable AI	Text Classification
Privacy-Preserving AI	SciPy	Text-to-Speech
Procedural Content Generation	Scikit-learn	Time Series Analysis
PyTorch	Seaborn	Topic Modeling
Q-Learning	Search Algorithms	Transfer Learning
Quantization	Self-Supervised Learning	Transferable Features
Quantum Machine Learning	Self-driving Cars	Transformer Models
Question Answering Systems	Semi-Supervised Learning	UMAP
ROC Curve	Sentiment Analysis	Underfitting
Random Forests	Sequence Modeling	Unsupervised Learning
Random Search	Simultaneous Localization and Mapping	User Behavior Analytics
Real-time Analytics	Smart Cities	Variational Autoencoders
Recall	Sound Synthesis	Video Analysis
Recommendation Systems	Spark	Video Object Tracking
Recurrent Neural Networks	Spatial Data Science	Virtual Reality and AI
Regularization	Speech Recognition	Voice Assistants
Reinforcement Learning	Speech Synthesis	Word2Vec
Relation Extraction	Streaming Data	Word Embeddings
Remote Sensing	StyleGAN	XGBoost
Responsible AI	Summarization Models	Zero-Shot Learning
Robotic Process Automation	Support Vector Machines	
Robotics Process Automation	Synthetic Data Generation	

### A.3. Web of Science

This section contains all the keywords found in the hottest articles regarding AI and Machine Learning on Web of Science.

1St	Acute Ischemic-Stroke	Ai-Based Interview
1St Law	Acute Lung	Air
1St Trimester	Acute Lymphoblastic-Leukemia	Air Co2 Flux
1St-Arrival Picking	Acute Myeloid-Leukemia	Air Pollution
1St-Line Therapy	Acute Myocardial	Air Quality Index
1St-Line Treatment	Acute Myocardial-Infarction	Air-Pollution
1St-Trimester	Acute Respiratory Syndrome	Air-Steam Gasification
2	Acute Tryptophan Depletion	Air-Temperature
2 Worlds	Adams-Bashforth	Airborne
3-D Inversion	Adaptation	Airborne Lidar
3-D Medical Watermarking	Adaptation Models	Airborne Transmission
3-Dimensional Stochastic-Analysis	Adaptation Strategies	Aisi
3D Asphalt Surfaces	Adaptive Cruise Control	Alcohol
3D Computer Vision	Adaptive Dynamic Programming	Algal Bloom
3D Object Recognition	Adaptive Mesh Refinement	Algorithm
3D Printing	Addiction	Algorithms
3D Shape Measurement	Addictive Behaviors	Alignment
3Rd Wave	Additive Manufacturing Of Pharmaceuticals	All-Cause
5G Security	Adenocarcinoma	Allele Diversity
5Th Generation Biosensor	Adenovirus-Vector Vaccine	Allen-Cahn Equation
617	Adjuvant	Allocare
6G	Adjuvant Chemotherapy	Allocation
6G Mobile Communication	Adjuvant Systemic Therapy	Allopatry
A-Beta	Administrative Data	Allowances Prices
Ab-Initio	Adolescence	Alphafold
Abandonment	Adolescent	Alpine Wetland
Abiotic Stress	Adolescents	Alspac
Abiotic Stresses	Ads-Cft Correspondence	Altered Flow Regimes
Aboveground Biomass	Adsorption	Alzheimer'S Disease
Abscisic-Acid	Aducanumab	Amazons Mechanical Turk
Absolute Error Mae	Adult Neurogenesis	Ambient Air-Pollution
Absorption Cross-Sections	Adults	American Society
Absorption Fine-Structure	Advanced Hepatocellular-Carcinoma	American Thoracic Society
Abundance	Advanced Oxidation	Amino Acids
Acc Expert Consensus Decision Pathway	Advanced Rectal-Cancer	Amino Sugars
Acc/Aha Clinical Practice Guidelines	Advection-Dominated Accretion	Ammonia
Acceleration	Adventitious Carbon	Amyloid Beta
Acceptance	Adversarial Transfer Network	Amyloid-Beta
Accessions	Adverse Events	Amylose Content
Accident Modification Factors	Aerosol Optical Depth	Analgesics
Accountability	Aerosol Transmission	Analysis Zone Level
Accumulation	Affordability	Analytic Hierarchy Process
Accuracy	Africa	Analytical
Accuracy Assessment	African	Analytical Models
Ace2	African Americans	Anchorage Mechanism
Acid Fractionation	Age	Andrology
Action Recognition	Aggregates	Angela Merkel
Activated Biochar	Aging	Angiotensin-Converting Enzyme
Active Galactic	Agricultural Green Total Factor Productivity	Animal-Sourced
Active Galactic Nuclei	Agricultural Insurance	Animals
Active Learning	Agriculture	Anion-Exchange-Membranes
Active Transport	Agxt Mutation	Anisotropic Power Spectrum
Active-Sites	Aha Scientific Statements	Annotation
Activity Indicators	Ai	Annotations
Activity Recognition	Ai Recruitment	Anomaly Detection
Acupuncture		Antagonist Oral Anticoagulants
Acute Coronary Events		Antarctic Piptoporus

Anti-Citrullinated Protein Antibodies	Augmentation	Biorefinery
Anti-Inflammatory	Augmented Reality	Biotechnology
Antibiotic Prescribing	Austerity	Bipolar Disorder
Antibiotic-Resistance	Authentication	Birth Rates
Antibiotic-Resistance Genes	Autism	Bismuth Oxychloride
Antibiotics	Autoencoder	Bitcoin
Antibodies	Autologous	Black People
Antibody	Automated Vehicles	Black Phosphorus
Anticancer Agents	Automatic Detection	Black-Box
Anticancer Peptides	Automatic Identification System	Black-Hole Binaries
Antidepressant Actions	Automation	Black-Hole Mergers
Antigen	Automation Bias	Black-White Disparities
Antigen-Dr Expression	Autonomation	Bleeding Risk-Assessment
Antimicrobial Resistance	Autonomous	Blockchain
Antimicrobial Stewardship	Autonomous Vehicles	Blockchains
Antimicrobial Use	Avatars	Blocking
Antimicrobials	Avena	Blood Lead Levels
Antioxidant Enzymes	Avoidable Healthcare System Cost	Blood-Brain-Barrier
Antiplatelet Therapy	B	Blood-Pressure
Antiracism	B-Containing Lipoproteins	Blueprint
Antiretroviral Therapy Initiation	Baby Bust	Bnt162B2
Antitumor-Activity	Bacillariophyceae	Bnt162B2 Vaccine
Antiviral Therapy	Bacillus-Subtilis	Board Diversity
Anus Cancer	Backorder	Bodies
Anxiety	Bacteria	Body Sensor Network
Anxiety Disorders	Bacterial	Body-Image
Aortic Stiffness	Bacterial Pneumonia	Body-Mass Index
Aphrs	Bad	Bond Coefficient
Apiaceae	Ban Gasoline Vehicles	Bone
Apiaceae Subfamily Apioideae	Band Achromatic Metalens	Bone-Marrow
Apobec3	Bariatric Surgery	Boosting
Apolipoprotein-E Genotype	Barley	Boosting Decision Trees
Apoptosis	Barrier	Bootstrap
Appendage	Baryon Acoustic-Oscillations	Bootstrap Rolling-Window Causality
Apple Detection	Basal Forebrain Atrophy	Bortezomib-Melphalan-Prednisone
Applicability Domain	Bat	Bounds
Applicants' Views	Batteries	Bp Neural Network
Application Decisions	Battery	Brain
Applications	Battery Life	Brain Fog
Approximate	Bayesian	Brain Structure
Approximation	Bayesian Networks	Brain Tumor
Aquaculture	Bayesian Regularized Bp Neural Network	Brain-Tumor Segmentation
Aquatic Contamination	Bayesian-Inference	Branch
Aquatic Environment	Bearing	Branch And Bound
Aqueous Media	Bearing Fault-Diagnosis	Brazil
Aqueous Zinc&#8208	Bee Colony Algorithm	Breast Cancer
Aqueous-Solution	Beef	Breast Neoplasms
Aqueous-Solutions	Beetles Coleoptera	Breastfeeding
Arabidopsis	Behavior	Breastmilk
Arbuscular Mycorrhizal Fungi	Behavior Change	Bridge
Architecture	Behavioral Insights	Bright End
Archive	Behavioral Reasoning Theory	Brite Hierarchical Classification
Archive Gsim	Belief	Broccoli
Area	Belief Space	Brominated Flame Retardants
Ariel3	Bells-Palsy	Brown-Rot Fungi
Artificial	Below-Ground	Bubbles
Artificial Bee Colony Optimization	Below-Ground Ecology	Building Change Detection
Artificial Intelligence	Benchmark	Buildings
Artificial Neural Network	Benchmark Testing	Built
Artificial Neural Networks	Benefits	Built Environment
Artificial Neural-Network	Besov Spaces	Burden
Artificial-Intelligence	Best Practice	Burnout
Artificial-Intelligence Ai	Best Practices	Business
As(Iii)	Bevacizumab	Business And Management
Aspect-Level Sentiment Analysis	Bf-Pso	Business Cycle
Assembly Process	Bias	Business Model Innovation
Assessment	Bibliometric Analysis	Buying-Shopping
Assisted	Bifidobacterium-Longum R0175	C-13-O-18
Association	Bifurcation	C-1S
Association Workgroups	Big Data	C-Reactive Protein
Astrocytes	Big Data Analytics	Cable Bolts
Astrometry	Big Data And Robots	Calcium-Oxalate Supersaturation
Astropy	Bilateral Cross-Modal Interaction	Calculator
Asymmetric Encoder-Decoder Model	Bilateral Network	Calibration
Asymptotic Distribution	Bile-Acids	Camels
Asynchronous Sensor	Binding Protein-1	Camera
At-Risk	Binding-Energy	Cameras
Athletes	Binding-Protein-Beta	Camrelizumab
Atmosphere	Binocular Stereo Vision	Canada
Atmospheric Fallout	Bio-Inspired Vision	Cancer
Atmospheric Methane	Bioavailability	Cancer Cases
Atomic Layer Deposition	Biochar	Cancer Diagnosis
Atomic Mass Evaluation	Biodegradation	Cancer Gene Expression
Atomic Mass Table	Biodiversity	Cancer Imaging
Atrial Fibrillation	Bioenergy	Cancer Incidence
Attack Detection	Bioinformatics	Cancer Mortality
Attacks	Biological	Cancer Statistics
Attention	Biological Soil Crusts	Canola
Attention Mechanism	Biomarker	Canonical Correlation-Analysis
Attention-Deficit/Hyperactivity Disorder	Biomarker Discovery	Canopy Height
Attitudes	Biomarkers	Canopy Structure
Attrition	Biomass	Capacity
Audio Watermarking	Biomaterial	Capecitabine
Auger-Electron-Spectroscopy	Biomedical Imaging	

Capital Accumulation	Chemical Inventory	Collection
Caputo	Chemical Pollution	Collections
Caputo-Fabrizio	Chemical-Composition	Collective Self-Esteem
Carbon	Chemical-Shifts	College-Students
Carbon Capture	Chemical-State	Collision-Induced Absorption
Carbon Cycle	Chemotherapy	Colon
Carbon Dioxide	Chemotropic Drug	Colon Cancer
Carbon Emission	Chengdu	Colon-Cancer
Carbon Emission Intensity	Chest Computed Tomography Image	Colonization
Carbon Emission Performance	Chest X-Ray	Color
Carbon Emission Reduction	Child Health	Colorectal Cancer
Carbon Emissions	Child & Adolescent Psychiatry	Colorectal-Cancer
Carbon Emissions Reduction	Childhood	Combination
Carbon Emission	Childhood Trauma	Combinatorial Optimization
Carbon Footprint	Childhood-Cancer	Combustion
Carbon Neutral	Children	Commodity Futures
Carbon Neutrality	China	Common Mental-Disorders
Carbon Peak	China Sea	Common-Spectrum Process
Carbon Price	China&#8217;	Communication
Carbon Pricing	Chloroquine Phosphate	Communication Systems
Carbon Reduction	Choice Architecture	Communication Technology
Carbon Sink	Cholesterol	Communication-Efficient
Carbon Trading Mechanism	Chronic Myelomonocytic Leukemia	Communities
Carbon-Based	Cik Cells	Community
Carbon-Density	Cinnamaldehyde	Community Ecology
Carbon-Dioxide Emissions	Circadian Clock	Community Health
Carbon-Isotope Discrimination	Circuit Theory	Community Wide Experiment
Carbon-Paste Electrode	Circuit-Theory	Comorbidity
Carbon-Reduction	Circular	Comparative
Carbon-Use Efficiency	Circular Economy	Comparative Ecology
Carbonate	Cisplatin	Comparison
Carbonemissionstradingmarket	Citation	Complete Chloroplast Genome
Carcinogenesis	Cities	Complex
Carcinoma	Citrullinated Peptide Antibodies	Complications
Cardiac-Resynchronization Therapy	Citrullus-Lanatus	Composite End-Points
Cardiovascular	City	Composites
Cardiovascular Benefits	Classification	Compositional Heterogeneity
Cardiovascular Disease	Classifiers	Compositional Rule Of
Cardiovascular Events	Clayff Molecular	Comprehension
Cardiovascular Magnetic-Resonance	Clearance	Compulsive Sexual-Behavior
Cardiovascular Risk	Climate	Computational
Cardiovascular Risk-Factors	Climate Change	Computational Chemistry
Cardiovascular-Disease Risk	Climate Change Issues	Computational Design
Care	Climate Feedbacks	Computational Modeling
Careers	Climate Models	Computational Physics
Carers	Climate Policy	Computed Tomography
Carotid Plaque	Climate Technologies	Computer
Carotid Restenosis	Climate-Change	Computer Architecture
Carp Cyprinus-Carpio	Clinical	Computer Security
Cas9	Clinical Characterization	Computer Vision
Case Definitions	Clinical Evidence	Computers
Case Volume	Clinical High-Risk	Concept Drift
Casp	Clinical Proteomics	Concrete
Catalase Activity	Clinical Text Classification	Concrete Crack
Catalog	Clinical-Features	Concurrent Chemoradiotherapy
Catalogs	Clinical-Outcomes	Conduct Disorder
Catalyst	Clinical-Trials	Conductivity
Catalysts	Clinimetric Properties	Conferences
Cationic Lipids	Cloud Computing	Confidence
Causal	Cloud Shadow	Confirmation
Causal Diagrams	Clumped Isotopes	Conformal And W Symmetry
Causal Inference	Cluster	Confusion Matrix
Cds	Clustering	Congestion
Ce(Iii) Ions	Clustering Algorithms	Conjugate Adsorbent
Ceemdan	Clustering Approach	Connected Vehicles
Cell	Cmpip6 Models	Connectivity
Cell Design	Cnn	Consensus
Cell Lines	Co 2 Emissions	Consensus Guidelines
Cell Lung-Cancer	Co-Benefits	Conservation
Cell Signaling	Co-Estimation	Consistent Basis-Sets
Cell-Cycle Arrest	Co-Infections	Consistent Nonparametric Test
Cell-Death	Co-Morbidity	Consolidation Therapy
Cells	Co2	Constitutive Model
Cellular Networks	Co2 Emission	Constraints
Cellular-Automaton	Co2 Emission Accounts	Constraints Created By Economic Growth
Central Brain Tumor Registry Of The	Co2 Emissions	Targets&Nbsps
United States	Co2 Emissions Performance	Consumer
Central Nervous System	Co3O4-Go Nanoparticles	Consumption
Central-Nervous-System	Coal	Consumption-Based Carbon Emissions
Cerebral Perfusion-Pressure	Cochran'S Q-Statistic	Consumption-Based Material Footprint
Cerebral-Cortex	Cognition	Consumption-Based Material Footprints
Cerebrospinal-Fluid	Cognitive Impairment	Contacts
Cervical-Cancer	Cognitive Reflection	Contamination
Cesium	Cognitive-Behavioral Therapy	Content Curation
Chain Fatty-Acids	Cognitive-Style	Conterminous United-States
Challenge	Cohort	Context Modeling
Challenges	Cohort Profile	Contextual
Change Detection	Cointegration	Continual Learning
Changing Profile	Coke	Contralateral Prophylactic Mastectomy
Chaperone-Mediated Autophagy	Cold-Plasma	Contrastive Learning
Charcoal Modified Adsorbent	Coleoptera	Contribute
Charge	Colitis	Conventional Semen Parameters
Checkpoint Inhibitors	Collaboration	Convergence
Chemical Footprint	Collaborative Work	Conversational Sentiment Analysis

Convolution	Cut	Dense Matter
Convolution Neural Network (Cnn)	Cyanobacterial Blooms	Density
Convolutional Network	Cyber-Physical Production	Density Functional Calculations
Convolutional Networks	Cyber-Physical Systems	Density Functional Theory
Convolutional Neural	Cyberattack	Density-Functional Theory
Convolutional Neural Network	Cybersecurity Industry	Dentate
Convolutional Neural Networks	Cycles Of Failure	Dependency Paths
Convolutional Neural Networks (Cnn)	Cystathione-Beta-Synthase	Depression
Convolutional Neural Networks (Cnns)	Cytokine Storm	Depression Symptoms
Convolutional Neural-Network	Cytotoxic Secondary Metabolites	Depression & Mood Disorders
Convolutional Neural-Networks	Czech Republic	Depressive
Cooking	D Deficiency	Depressive Symptoms
Cooling Energy Trends	D-Aspartate Antagonist	Depth
Cooperative Coevolution	Dabrafenib	Depth Estimation
Copd	Daily Activity Recognition	Depth Functions
Coping	Damage	Depthwise Separable Convolution
Copy Number Variation	Damage Evolution	Der-Waals
Cord Blood	Dammed Lake	Descent
Coronary-Artery Dissection	Dark Energy	Desert Steppe
Coronary-Heart-Disease	Dark Matter	Design
Coronavirus	Dark Matter Phenomenology	Design Methodology
Coronavirus Disease 2019	Dark Radiation	Detection Capturing And Recovery
Coronavirus Infection	Data	Detectors
Corporate	Data Approximation Scheme	Determinants
Corporate Environmental Investment	Data Assimilation	Dft
Corporate Governance Research	Data Assimilation Method	Diabetes
Corporate High-Quality Development	Data Envelopment	Diabetes Mellitus
Corporate Resilience	Data Envelopment Analysis	Diabetic Retinopathy
Corporate Social Responsibility	Data Fusion	Diagnosed Multiple-Myeloma
Corporate Social-Responsibility	Data Mining	Diagnosis
Correlate	Data Models	Diagnostic Errors
Correlation	Data Quality	Diagnostic Guidelines
Correlation Analysis	Data Set	Dialysis
Corticotropin-Releasing Hormone	Data Set Limits	Did Model
Cosmological Parameter	Data-Driven	Dielectric Metasurfaces
Cosmological Parameters	Data-Driven Control	Diet
Cosmology	Database	Diet-Induced Obesity
Cost-Benefit Analysis	Databases	Dietary Factors
Cost-Effectiveness Analyses	Dataset	Difference Vegetation Index
Cost-Effectiveness Analysis	Dbcan	Differences-In-Differences
Costs	Dce-Mri	Differential
Count Data	De Ficiency	Differential Abundance Analysis
Countries	De Novo Assembly	Differential Equations
Country-Of-Origin	Dea Model	Differential Evolution
Coupled-Cluster	Deal.Ii	Differential Expression
Coupling And Coordination	Death Rates	Differential Expression Analysis
Coupling Coordination Degree (Ccd)	Debonding Behaviour	Differentiated Thyroid-Cancer
Covariance Based Structural Equation	Debris	Differentiation Therapy
Covariance Estimation	Decay	Diffusion
Cover Change	Decentralized Microgrids	Digital
Coverage	Decision	Digital Agriculture
Covid	Decision Making	Digital Communication Technologies
Covid-19	Decision Making And	Digital Economy
Covid-19 Outbreak	Decision Trees	Digital Finance
Covid-19 Pandemic	Decision-Making	Digital Footprint
Covid-19 Vaccine	Decision-Tree Classification	Digital Holographic Microscopy
Covid19	Decisions	Digital Media Literacy
Cox Regression	Decisive Role	Digital Photography
Cr-Fe-Ni	Decoding	Digital Sky Survey
Crack	Deep	Digital Technology
Crash Frequency	Deep Active Learning	Digital Transformation
Crash Type	Deep Belief Networks	Digital Twin
Credit Constraints	Deep Ensembles	Digital Twins
Credit Policy	Deep Feature Fusion	Dimensionality
Crispr	Deep Knowledge Tracing	Dinophyceae
Criteria	Deep Learning	Diode-Laser
Critical Situational	Deep Learning Framework	Dioxide
Critical-Path-Analysis	Deep Neural	Dioxide Emissions
Critically-Ill Patients	Deep Neural Network	Direct Electron-Transfer
Crohn'S Disease	Deep Neural Networks	Direct Numerical-Simulation
Crohns-Disease	Deep Neural-Network	Directed Acyclic Graphs
Crop Productivity	Deep Neural-Networks	Direction Information
Cross Modality	Deep Translation	Disaster
Cross-Lagged Panel Models	Deep-Water Access	Discharge
Cross-Layer Interaction	Deepxde	Discovery
Cross-Match	Defective Uio-67/Go	Discrete
Cross-Sectional	Defense	Discrete Wavelet Transform
Cross-Sectional Studies	Defense Response	Discrimination
Cross-Validation	Deficits	Disease
Crude Oil Market	Definition	Disease Control
Crude-Oil	Deforestation	Disease Outbreaks
Crude-Oil Price	Deformation	Disease Resistance
Crushed	Deformation Mechanism	Disease Severity
Cryptocurrency	Degradation	Diseases
Crystal-Structure	Delay	Disorder
Crystals	Delays	Disorders
Cso	Delivery	Disparities
Cspbb3 Nanocrystals	Delta Variant Predominance	Dissipative Multiconstituent Medium
Csr	Delta(47) Analysis	Dissolved Organic-Matter
Ct	Demand	Dissolved-Gas Analysis
Cu	Demands	Dissolved-Oxygen
Cultivated Land Protection	Dementia	Distress
Curse	Democracy	Distributed Photovoltaic
Customer	Dendritic Cells	

Diversification	Effort	Environmental Kuznets Curve
Diversity	Ejection Fraction	Environmental Performance
Dmusp-Ols	Ekc	Environmental Policy
Dna	Ekc Hypothesis	Environmental Pollution
Dna Barcoding	Elastic	Environmental Protection Tax
Dna Shape-Features	Elastic Mechanics	Environmental Quality
Dna-Binding	Electric Vehicle-Batteries	Environmental Regulation
Dna-Repair Defect	Electric Vehicles	Environmental Sustainability
Domain	Electric-Power Consumption	Environmental Tax
Domain Adaptation	Electrocardiography	Environmental-Policy Stringency
Domestication	Electrocatalysts	Environmental-Pollution
Dominant Male-Sterility	Electrocatalytic	Environmental-Quality
Donation	Electrocatalytic Reduction	Environmental-Regulation
Doped Porous Carbon	Electrochemical Capacitors	Environmental-Regulations
Double Dividend Hypothesis	Electrochemical Energy-Storage	Environments
Double-Beta-Decay	Electrochemical Impedance Spectroscopy	Enzymes
Double-Blind	Electrochemical Model	Epidemiology
Double-Strand Breaks	Electrochemical Sensor	Epigenetic Inheritance
Driven	Electron Correlation	Epistemic Exclusion
Drivers	Electronic	Epistemic Uncertainty
Driving Forces	Electronic Nose	Epithelial-Mesenchymal Transition
Drones	Electronic Properties	Equation
Dropout	Electrons	Equation-Of-State
Drought	Electrophysiology	Equations
Drought Stress	Elements	Equity
Drought Tolerance	Elevated Neutrophil	Erosion Rates
Drought-Induced Mortality	Elliptic	Error-Correction
Drug Resistance	Elliptic-Equations	Escape
Drug-Delivery System	Emden	Escherichia-Coli
Dual-Phase Xenon	Emergency-Department Visits	Essential Medicines
Durable Dominance	Emerging	Ester Biosynthesis
Durvalumab	Emerging Contaminants	Estimated Annual
Dutch Disease	Emerging Infections	Estimation
Dye Adsorption	Emerging Seven	Estimation Algorithm
Dyes	Emission	Ethnic-Identity
Dynamic	Emission Peak	Ethnicity
Dynamic Histogram Equalization	Emission Reduction	Ethylene
Dynamic Immersion	Emission Reductions	Ethylene-Inducing Xylanase
Dynamic Process	Emissions	Etoposide
Dynamic Regional Homogeneity	Emotion Recognition	European Union
Dynamic Spillover	Emotional Recurrent Unit	European-Society
Dynamic Threshold Model	Empirical Mode Decomposition	Evaluation
Dynamic-Analysis	Empirical Risk Minimization	Evasion
Dynamic-Range	Empirical Test	Event Cameras
Dynamics	Empirical-Evidence	Everolimus
Dynamics Analysis	Empirical-Research	Evidence-Based Medicine
Dyspnea	Employee Welfare	Evidence-Based Recommendations
Early Season Mapping	Employment	Evolution
Early Yield Estimation	Ems	Evolution-Equations
Earth Observation	Enabling	Exchange
Earth-Sized Planet	Enabling Technologies	Exchanger
Easterlin Paradox	Encephalitis	Excitations
Eastern Mediterranean	Encryption	Exercise
Eastern-Europe	Encyclopedia	Exergy Analysis
Eating Disorders	End Collision Risks	Exome Sequencing
Eco-Friendly Concrete	End-To-End Network	Exoplanets
Eco-Innovation	Endangered Languages	Expansion
Ecological	Endocrine Disrupting	Experimental-Design
Ecological Economics	Endometrial	Expert Knowledge
Ecological Footprint	Endoplasmic-Reticulum Stress	Explainable Ai
Ecological Risk Assessment	Endothelial Dysfunction	Explainable Artificial Intelligence (Xai)
Ecological Security	Endovascular Treatment	Explanations
Ecological Sustainability	Energy	Exploration
Ecology	Energy Consumption	Exploration Geophysics
Econometrics	Energy Dissipation Sequences	Exploratory Data Analysis
Economic	Energy Efficiency	Explosion
Economic Burden	Energy Efficiency Gap	Export
Economic Complexity	Energy Endowment	Exposure
Economic Evaluation	Energy Harvesting	Exposure Assessment
Economic Growth	Energy Policy	Expression
Economic Growth Target	Energy Saving And Emission Reduction	Extended Marine Predators Algorithm
Economic Policy Uncertainty	Efficiency	Extension Statement
Economic Production Quantity	Energy Statistics Revisions	Extinction Risk
Economic-Development	Energy Structure	Extracellular Dopamine
Economic-Growth	Energy Systems	Extract Characterisation
Economics	Energy Transition	Extraction
Economy	Energy-Consumption	Extragalactic Distance Database
Ecosystem Services	Energy-Environmental	Extremal Optimization
Ecotoxicity	England	Extreme
Eddy Covariance Measurements	Enhanced Infrared-Spectroscopy	Extreme Gradient Boosting
Edge Centrality	Enhanced Synaptic Plasticity	Extreme Learning Machine
Edge Computing	Enough	Extreme Learning-Machine
Edge-Califa	Enrichment Analysis	Eye
Education	Ensembles	Fabrication
Education Software	Enso Amplitude Change	Face
Educational-Attainment	Enteral Nutrition	Face Recognition
Eeg	Entrepreneurial Futility	Face Super-Resolution
Effect Sizes	Entropy	Facial Characteristics
Effect Traits	Environment	Facial Feminization Surgery
Efficacy	Environmental	Facile Synthesis
Efficiency	Environmental Degradation	Faculty Of Color
Efficient	Environmental Efficiency	Faecalibacterium-Prausnitzii
Efficient Selenium(IV) Detection	Environmental Impacts	Fairness
Efficientnet	Environmental Kuznets	Fairness And Bias In Artificial Intelligence

Fake News	Formation Rates	Generative Model
Family	Fossil-Fuel Co2	Genetic
Family Classification	Foundation	Genetic Algorithm
Family-Work Conflict	Founding-Family Ownership	Genetic Algorithms
Fast Radio Burst	Fp	Genetic Susceptibility
Faster R-Cnn	Fractal-Fractional Operator	Genetic-Variants
Fat Distribution	Fracture	Genome
Fatigue	Fragility Analysis	Genome Annotation
Fatty-Acid-Composition	Framework	Genome Database
Fault Detection	Frbr	Genome Sequencing
Fault Diagnosis	Free Energy	Genome-Wide Association
Fault-Diagnosis	Free-Vibration Analysis	Genome-Wide Expression
Fda Approval	Freeze-Thaw Cycle	Genomic
Feature	Freeze-Thaw Cycles	Genomic Characterization
Feature Extraction	Freezing Tolerance	Genomic Surveillance
Feature Fusion	Freight Truck-Related Crashes	Genotype
Feature Selection	Frequency	Genotype Imputation
Feature Similarity	Frequent Impact Loads	Genus
Feature-Extraction	Fresh	Geodetector
Feature-Selection	Fresh Agricultural Products	Geostatistical Analysis
Features	Frontotemporal Lobar Degeneration	Germinal Center
Fecal Immunochemical Test	Fruit Detection	Ghq
Fecal Metabolome	Fruit Flavor	Giant Molecular Clouds
Fecal Microbiota	Fruit-Development	Ginzburg-Landau Model
Federated Learning (Fl)	Fuel-Cell Cathodes	Gis
Feedback	Full Implication Algorithms	Glaucoma
Ferromagnetic	Fully	Glioblastoma Multiforme
Fertility	Fully Convolutional Networks	Global
Fetal Testosterone	Functional	Global Burden
Fg Nanoplate	Functional Architecture	Global Burden Of Disease
Fgk	Functional Connectivity	Global Change
Fiber	Functional Genomics	Global Co2 Emissions
Fibrillary Acidic Protein	Functional Time-Series	Global Estimates
Field	Functional Traits	Global Level
Field-Effect Transistors	Functionally Graded Sandwich	Global Optimization
Films	Functionals	Global Surveillance
Filtration	Fungal Pathogens	Global Value Chain
Financial	Fungus &Ndash;	Global Vegetation Models
Financial Development	Fusion	Globalization
Financial Globalization	Future	Globally
Financial Inclusion	Future Cities	Globalsoilmap
Financial Performance	Fuzzy Inference System	Globocan
Financial Risk	Fuzzy Logic	Globular Clusters: General
Financial Sustainability	Fuzzy Reasoning	Glomerular-Filtration-Rate
Financing Constraints	Fuzzy Set	Gnss
Fine Aggregate	Fuzzy-Logic	Goat Meat
Fine Particulate Matter	Fuzzy-Systems	Gold Standard Validation
Fine-Grained Object Detection And Recognition	G7 Countries	Gonadotropin-Releasing-Hormone
Fine-Particulate Matter	Galaxies	Governance
Fine-Structure Constant	Galaxy Clusters	Government Ideology
Finite Elements	Galaxy Formation	Government Intervention
Finite-Element-Method	Galaxy:	Government Subsidy
Finite-Size	Games	Gpr
Fintech	Gap-Free Genome	Gradient
Fintech Innovation	Gas	Gradients
Firm	Gas Depletion Time	Gradual Did
Firm Performance	Gas Sensing	Grain-Yield
First -Order Convolution	Gas Sensors	Grandmothers
First Quality	Gas-Chromatography	Granite
Fiscal	Gasification	Granular Flow
Fiscal Incentives	Gastric Bypass-Surgery	Granule Cells
Fish Communities	Gastritis	Graph
Fish Meal	Gated Recurrent Units (Grus)	Graph Convolutional
Five-Dimension Model	Gaussian Process Emulator	Graph Data
Fixed-Bed Column	Gaussian Process Regression	Graph Databases
Flare	Gaussian-Basis Sets	Graph Neural Network
Flavor	Gbsi	Graph Query Languages
Flocculation	Gedi	Graph Reasoning
Flood Prediction	Gemcitabine	Graphene
Flow	Gender	Graphene Oxide
Flow Prediction	Gender Diversity	Graphene Oxide Sheets
Flow Reactor	Gene	Graphical Model Theory
Flow Regimes	Gene Array	Graphical User
Fluid Flow	Gene Discovery	Graphics Processing Units
Fluvoxamine	Gene Expression	Graphite Oxide
Fmri	Gene Expression Data	Grass
Foam	Gene Pav	Gravitational Search Algorithm
Focus Group	Gene Selection	Gravitational Waves
Folic-Acid	Gene-Cluster	Gravitational-Waves
Follow-Up	Gene-Expression	Gravity-Waves
Fomitopsidaceae	General Framework	Green
Food	General-Population	Green Credit
Food Security	General-Surgery	Green Development
Footprint	Generalist Microbiota	Green Economic Growth
For-The-Ligule	Generalizability	Green Economies
Force	Generalization Error	Green Economy
Forced Expiratory Volume	Generalized	Green Energy
Forecast	Generation	Green Energy Index
Forecasting	Generation Scotland	Green Finance
Foreground Maps	Generative	Green Financing
Foreign Direct-Investment	Generative Adversarial Network	Green Innovation
Forest	Generative Adversarial Networks	Green Innovation Efficiency
Forest Fire	Generative Biology	Green Investment

Green Process Innovation	Herd-Immunity	Illegal Cultivated Land Use
Green Product Innovation	Heritability	Illegal Farmland Conversion
Green Productivity	Heterogeneity	Illness
Green Public Finance	Heterojunction Photocatalyst	Illumination
Green Recovery Scheme	Heuristic Algorithms	Image
Green Tea Polyphenol	Hexagonal	Image Analysis
Green Technological Innovation	Hf	Image Authentication
Green Technology	Hierarchical Representation	Image Classification
Green Technology Innovations	High Adsorption	Image Detection
Green Total Factor Energy Efficiency	High-Flow Oxygen	Image Fusion
Green Total Factor Productivity	High-Mountain Asia	Image Quality
Greenhouse Gas	High-Performance	Image Quantitation
Greenhouse Gas Emissions	High-Precision Measurements	Image Resolution
Greenhouse-Gas	High-Precision Spindles	Image Restoration
Greenhouse-Gas Emissions	High-Quality Economic Development	Image Segmentation
Gross Primary Production	High-Redshift Galaxies	Image Superresolution
Ground	High-Resolution	Image Thinning
Ground Biomass	High-Resolution Aerial	Image Transmission
Ground Motions	High-Reynolds-Number	Image-Analysis
Grounded Theory	High-Risk	Imagery
Groundwater	High-Speed Rail	Images
Groundwater Remediation	High-Spin States	Imaging
Grouted Rock Bolt	High-Tech Industries	Imaging Genetics
Growth	High-Temperature	Imbalanced
Growth-Performance	High-Temperature Oxidation	Imcg100
Growth-Promoting Rhizobacteria	High-Throughput	Immense
Guangzhou	High&#8208;	Immune Cells
Gudermannian Neural Network	Higher Education	Immune Checkpoint Blockade
Guidance	Higher Order	Immune Checkpoint Inhibitor
Guidelines	Higher-Education	Immune Response
Gut Microbiome	Highly Efficient	Immune Responses
Gut Microbiota	Hilbert Space	Immune-Related Adverse Event
Gut-Lung Axis	Histone Deacetylase Inhibitors	Immune-Related Genes
Gwas	Hitran	Immunity
Gymnodiniales	Hiv	Immunization
H Alpha	Hla System	Immunogenicity
H1N1	Holographic Dark Energy	Immunotherapy
H5N1	Homozygous Familial	Impact
Half-Life	Hospital-On-Chip	Impacts
Halide Perovskite Nanocrystals	Hospitals	Implementation
Hall-Petch Relationship	Host	Implementation Framework
Hallmarks	Host Mitochondria	Implementation Research
Hallucinating Faces	Host-Specificity	Implementation Science
Hammerstein-Wiener Model	Hot Weather	Improved Loss Function
Hand Osteoarthritis	House Photos	Improving
Handbook	House Price Appreciation Rate	Improving Accuracy
Happiness	Household-Owned Farm Machinery	Impulsive Waves
Hardened	Housing Price	In Vitro
Hardened	Hubble Constant	In-Situ
Hardware	Hubble Frontier Fields	In-Vitro
Harmful	Hubble Tension	In-Vivo
Harmonic Vibrational Frequencies	Human	Incentive Environmental Regulations
Hawks Optimization Algorithm	Human Activity Recognition	Incentives
Hazardous Metals Adsorption	Human Biomonitoring	Incidence
Hazardous Pollutants	Human Breastmilk	Incidental Findings
Head	Human Capital	Incision Cataract-Surgery
Headache	Human Coronaviruses	Including Microplastics
Headwater Streams	Human Health	Income Countries
Health	Human Papillomavirus (Hpv)	Income Diversification
Health Behaviour	Human Reproduction	Income Group Level
Health Care Workers	Human Whole Blood	Income Inequality
Health Disparities	Human-Papillomavirus Infection	Increase Suicide Rates
Health Economics	Hunger Games Search	Independent Component Analysis
Health Effects	Hybrid	India
Health Estimation	Hybrid Nanofluid	Indian-Summer Monsoon
Health Professional	Hydraulic Conductivity	Indicators
Health Service Research	Hydraulic Servo	Indices
Health-Care	Hydride	Individual Effects
Health-Care Interventions	Hydro Energy	Individual Tree Crown
Health-Care Workers	Hydrodynamic Journal Bearing	Individual-Differences
Health-Care-System	Hydrogen	Individuals
Health-Risks	Hydrogen Bonds	Induced Myopathy
Healthcare	Hydrogen Evolution	Industrial Application
Heart	Hydrogen Fluoride	Industrial Internet Of Things
Heart Failure	Hydrogen Purification	Industrial Revolution 4
Heart Failure With	Hydrogen Sulfide	Industrial Structure Upgrading
Heart Failure With Reduced	Hydrogeological Model	Industrialization
Heat	Hydrological Observation Reports	Industry
Heat Recovery	Hydropower Policy	Industry 4
Heat Transfer	Hydrothermal	Industry 4.0
Heat-Related Mortality	Hydroxyl Radical Oh	Industry 5.0
Heat-Stress	Hyperglycemia	Inequalities
Heat-Transfer	Hyperparameter	Inequality
Heating Systems	Hypersensitivity Reactions	Inertial
Heavy Metal	Hyperspectral	Infant Formulas
Heavy Metal Contamination	Hyperspectral Image Classification	Infants'
Heavy-Metal Ions	Hyperspectral Imaging	Infarct Size
Hebei	Hypertension	Infection
Heightmap	Ibrutinib	Infectious Disease
Helicobacter	Icd-11	Infectious Diseases
Helicobacter Pylori	Idarubicin	Inference
Hematopoietic-Cell Transplantation	Identification	Inflammasome Activation
Hemodialysis	Identifying	Inflammation
Hepatic Resection	Identity	Inflammatory Response
Hepatitis		

Inflammatory-Bowel-Disease	Iot Architecture	Left-Ventricular Dysfunction
Influencing	Ipilimumab	Left-Ventricular Hypertrophy
Influencing Factors	Iron	Length
Influenza	Iron(Iii) Complexes	Leptin Receptor
Influenza Vaccine	IrT	Leptin Resistance
Influenza-A H1N1	Ischemia/Reperfusion	Leptin-Based Therapies
Influenza-Virus	Ischemic-Stroke	Lexicalized Dependency Paths
Information	Ising Models	Lexicon-Based
Information Communication Technology	Isotherm	Liao River
Information Fusion	Iteration Method	Lidar
Information-Content	Iterative Subsampling	Lidar Data
Information-System Promis(R)	Java Interface	Life
Informed	Jet-Stirred Reactor	Life-Cycle Assessment
Infrared Patch (Ipi) Image	Job Productivity	Life-Cycle Management
Infrared Search And (Irst) Track System	Job Resources	Life-Span
Infrared-Emission	Joint Hbm4Eu Survey	Life-Style Intervention
Infrastructure	Joint Source-Channel Coding	Light
Inhba Gene-Expression	K-Complex	Light-Emitting-Diodes
Inhibition	K-Complexes	Light-Induced Thermoelastic
Inhibitors	Kaplan-Meier Plot	Lightgbm
Injury	Karst	Lighting
Injury Severity	Karst Mining Area	Lignin
Innate Immunity	Kegg	Lignocellulosic Biomass
Innovation	Kegg Mapper	Limb-Darkening Coefficients
Innovation Activities	Keller-Segel System	Limited Replicability
Innovation In Environmental-Related Technologies	Kernel	Lineage
Inoculation	Keywords	Lineage Plasticity
Input-Output-Analysis	Kidney Cancer	Link
Inrush	Kidney Transplant	Linking
Insecurity	Kin	Lipid Accumulation Product Index
Insights	Kinase	Lipid Nanoparticles
Institutional Quality	Kinetic Model	Lipidomics
Instrument Strength	Kinetics	Lipids
Instrument Validity	Kiwifruit	Lipoprotein Remnants
Instrumental Variable Analysis	Knee Osteoarthritis	Liquid Biopsy
Instrumental Variables	Knowledge	Liquid-Chromatography
Integrated	Knowledge Acquisition	Liquid-Solid
Integrated Information-Theory	Knowledge Based Systems	Liraglutide
Integrating	Knowledge Distillation	Lisflood-Fp
Integration	Knowledge Engineering	List
Integration Factor Methods	Knowledge Graphs	Literacy
Integrity	Koh Activation	Lithium-Ion Batteries
Intelligence	Kuznets Curve	Lithium-Ion Battery
Intelligent Fault Diagnosis	Kuznets Curve Hypothesis	Lithium-Metal Batteries
Intelligent Sports	L. Essential Oil	Litter
Intensity	L452R	Liver Metastasis
Interaction Network	Labels	Liver-Abscess
Interactive Effects	Labor	Liver-Cancer
Interlaboratory Calibration	Laboratory Transition-Probabilities	Lnp
Intermediary Effect Model	Laccaria-Bicolor	Load
Intermittent Rivers	Lactobacillus-Helveticus R0052	Load Balancing
Internal-Friction	Lagrange Multiplier Test	Local
International Crude-Oil	Laminar Burning Velocities	Local Food
International-Trade	Lammps	Local Presence
Internet	Land	Localization
Internet Development	Land Cover	Locally Advanced Nsclc
Internet Gaming Disorder	Land Degradation	Location Awareness
Internet Of	Land Sustainability	Loci
Internet Of Medical Things	Land Use	Lockdown
Internet Of Things	Land-Cover	Loess Landslide
Internet Of Things (Iot)	Land-Cover Change	Logistic-Regression
Interpretability	Land-Cover Classification	Logistic-Regression Models
Interpretation	Land-Use	Loneliness
Interpreting Effect Sizes	Land-Use Change	Long
Interstellar Dust	Land-Use Simulation	Long Covid
Interstitial Fluid	Land-Use-Change	Long Covid-19 Symptoms
Interventions	Landsat	Long-Covid
Interviews	Landscape	Long-Covid Symptoms
Intestinal	Landslide Susceptibility Mapping	Long-Haul
Intestinal Microbiome	Landslide Susceptibility Prediction	Long-Read Sequencing
Intestinal Microbiota	Lanthanum	Long-Term Exposure
Intestine	Large-Eddy Simulation	Long-Term Trends
Intracerebral Hemorrhage	Large-Magellanic-Cloud	Longitudinal
Intravenous Ketamine	Large-Scale	Longitudinal Data
Intrinsic Disorder	Large-Scale Dam	Longitudinal Research
Intrinsic Pontine Glioma	Laser Radar	Low Cost Materials
Intron	Lasso	Low Latency
Intrusion	Late	Low Self-Esteem
Intrusion Detection	Latency	Low Temperature Plasma
Intuitionistic Fuzzy Entropy	Latent Variable Model	Low- And Middle -Income Countries
Inventory	Lattice Oxygen	Low-Birth-Weight
Inverse Problems	Lattice Vibrations	Low-Carbon Cities
Inversion	Layered Double Hydroxide	Low-Carbon Development
Investment	Ld Score Regression	Low-Carbon Pilot
Investment Evidence	Lda Topic Model	Low-Density Lipoproteins
Investments	Ldl Cholesterol	Low-Frequency
Io Model	Lead-Free Double Perovskites	Low-Head Dam
Ion Batteries	Leadership	Low-Rank
Ion Battery	Leaf Hydraulic Vulnerability	Low-Temperature
Ion Migration	Leaf-Area Index	Lstm
Ionizable Lipid	Learning	Lubrication Mechanism
Ionosphere	Learning Loss	Lucc Simulation
IoT	Learning Systems	Lung
	Least Square Support Vector Machine	Lung Transplantation

Lung-Function	Medical Image	Molecular
Luteinizing-Hormone	Medical Image Segmentation	Molecular Clouds
Lyman Break Galaxies	Medical Services	Molecular Dynamics
Lymph-Node	Medical-Treatment	Molecular Gas Properties
Lymphoma	Medicinal Plant	Molecular Hybridization
Machine	Mek	Molecular Markers
Machine -Learning Algorithms	Memory	Molecular Mechanism
Machine Learning	Memory Management	Molecular Pathological Epidemiology
Machine Learning Algorithm	Mendelian Randomisation	Molecular Simulation
Machine Learning Algorithms	Mental Disorder	Molecular Spectroscopy
Machine Learning Models	Mental Disorders	Molecular-Vibration
Machine Learning-Methods	Mental Health	Molecular-Data Reveal
Machine Learning-Model	Mental Health Problems	Molecular-Dynamics
Machine Sense Of Smell	Mental Illness	Molecular-Dynamics Simulations
Machine Tools	Mental-Health	Molnupiravir
Machine Vision	Mental-Health Problems	Monitoring
Machine-Learning (Ml)	Mental-Health-Services	Monkeypox Versus Covid-19
Machine-Tools	Messenger-Rna Vaccine	Monkeypox Worries
Machinery	Meta Analyses	Morbidity
Macroecology	Meta-Analysis	Moringa Oleifera
Macroinvertebrate Communities	Metaanalyses	Morphological
Macular Degeneration	Metabotropic Glutamate-Receptor	Mortality
Macular Edema	Metagenomics	Mos2
Mad	Metabolic-Activities	Motif Database
Magnetar	Metabolism	Motor Areas
Magnetic	Metabolome	Motor-Vehicle Crashes
Magnetic Fluids	Metabolomics Data	Mount-Pinatubo
Magnetic Resonance Imaging	Metabotropic Glutamate-Receptor	Movement-Disorders
Magnetic Resonance Imaging (Mri)	Metagenomics	Moving Bed Heat Exchanger
Magnetic-Field	Metaheuristic	Mri
Magnetic-Field Structure	Metaheuristics	Mrio Model
Magnetic-Resonance-Spectroscopy	Metal-Based Nanocomposite	Mrna
Magnetism	Metal-Organic Framework	Mrna Vaccine
Magnetization	Methicillin-Resistant	Ms
Magnonics	Methodology	Multi
Magnons	Methods	Multi -Criteria Decision Making
Maharashtra	Methods: Data Analysis	Multi -Scale Feature
Main Galaxy	Methods: Observational	Multi-Cancer Early Detection
Main-Sequence	Methylation	Multi-Domain Information Fusion
Major	Metrics	Multi-Focus Image
Major Depression	Microbial	Multi-Gene Phylogeny
Major Depressive Disorder	Microbial Activity	Multi-Modal
Male Infertility	Microbial Communities	Multi-Modal Multi-Objective
Male Reproductive Disorders	Microbial Community	Multi-Modality
Maltreatment	Microbial Community Abundance	Multi-Mode Data Augmentation
Management	Microbial Diversity	Multi-Omics
Mandatory Environmental	Microbial Remediation	Multi-Scale Heterogeneous
Maneuvering Targets	Microbiome	Multi-Task Learning
Manufacture Of Medicinal	Microeukaryotic Plankton	Multicenter
Manufacturing System Design	Microfluidics	Multicriteria Decision-Making
Maom	Microorganisms	Multilevel Methods
Map	Microplastic	Multimodal Remote Sensing
Mapping	Microplastics	Multiple
Maps	Microscopy	Multiple Description
Marine-Environment	Microservice Architectures	Multiple System Atrophy
Market	Microwave-Assisted Pyrolysis	Multiple Tasks
Market Indexes	Middle	Multiresolution Analysis
Market Research	Middle East	Multiscale
Market Volatility	Middle-Income	Multispecies Coalescent
Marketization	Middle-Income Countries	Multitarget
Mass	Migraine	Multitasking
Mass Spectrometry	Mild	Multivariable Mendelian
Mass-Distribution	Milky-Way	Multivariate
Mass-Spectrometry	Mine Water	Mung Bean
Massive	Mineral	Municipal Solid-Waste
Massive Data	Mini-Mental-State	Mutant
Massive Galaxies	Minimal Residual	Mutant-Cells
Matching	Minimum	Mutation
Materials Modeling	Mining Soil	Mutation Library
Maternal Health	Mirrors	Mutations
Maternal Support	Missing Data	Mutton
Mathematical Model	Missing Labels	Mxenes
Mathematical Model Of Drinking	Mitigation	N Shaped
Mathematical Modeling	Mitigation Targets	Nac
Matrix	Mixed Carbonate/Lipf6 Electrolyte	Naive Patients
Matrix Factorization	Mixed Reality	Named Entity Recognition
Matrix Product States	Mixture	Nanoparticles
Matter	Mlr	Nanoplastics
Maximum Bound Principle	Mm	Nanoplastics
Maximum Clique	Mobile	Nanotechnology
Mdr/Xdr-Tb	Mobile Harvesting Robots	Narrative Review
Measurement	Mobile Lidar	National Institute
Meat	Mobile Phone Use	National-Health
Mechanical	Mode	Nativity Status
Mechanical Properties	Model	Natriuretic Peptides
Mechanical Ventilation	Model Compression	Natural Gas Sweetening
Mechanism	Modeling	Natural History Collections
Mechanisms	Modelling	Natural Language Processing
Media	Models	Natural Resource
Mediating Effect	Modifiable Risk-Factors	Natural Resource Rents
Mediation	Modified Argan	Natural Resources
Mediation Analysis	Modified Chaplygin-Gas	Natural Waters
Medical Diagnostic Imaging	Modulation	Natural-Convection
Medical Education	Module	Natural-History

Natural-Resource Dependence	Nrp1	Oxidation
Ncov-19 Azd1222 Vaccine	Nsclc	Oxidative Stress
Near-Fault Pulse Seismic	Nuclear	Oxygen
Near-Infrared Flares	Nuclear Energy	Oxygen Reduction
Necrosis-Factor-Alpha	Nuclear Medicine	Oxygen Species Homeostasis
Neighborhood Environmental Factors	Nucleic-Acid Detection	Oxygen Vacancies
Neighborhoods	Nucleus Segmentation	Oxygen-Evolution Reaction
Neisseria-Meningitidis	Nudge	Oxygen-Glucose Deprivation
Neonatal Intensive Care Unit	Number	P.1
Neonate	Numerical	Pacific
Nephrectomy	Numerical Simulation	Package
Nest Connection	Numerical Weather Prediction	Packaging Signal
Net	Numerical-Simulation	Packet Loss
Net Ecosystem Co2	Nurses	Pain Care Equity
Net Primary Productivity	Nutrient Release	Pain Disparities
Netherlands	Nutrition	Pan-Genome
Network	Nutritional	Pan-Sharpening Method
Network Architecture	O2O Channeling	Pandemic
Network Psychometrics	Obese	Pandemics
Network Security	Obesity	Panel
Networks	Object	Panel-Data
Neural	Object Detection	Panel-Data Analysis
Neural Networks	Object Detection And Recognition	Parabolic-Elliptic System
Neural-Network	Object Tracking	Paradigm
Neural-Network Ann	Objective Deployment Optimization	Parallaxes
Neural-Network Model	Observational Cohort	Parallel
Neural-Network Prediction	Obsessive-Compulsive Disorder	Parallel Algorithms
Neural-Networks	Obstructive	Parameters
Neuro-Covid-19	Obstructive Pulmonary-Disease	Parametrized Pdes
Neurodegeneration	Occult Blood-Test	Paraquat
Neuroendocrine Tumors	Ocean Personality Model	Parietal Epithelial-Cells
Neurological	Of-Charge	Parkinson'S Disease
Neurological Manifestations	Of-Health	Parkinsons-Disease
Neurological Symptoms	Of-The-Art	Partial Least-Squares
Neuromuscular Electrical-Stimulation	Ofdi	Partial-Differential-Equations
Neuronal Apoptosis	Off-Label Use	Participants
Neurons	Oil	Participation
Neuropsychological Assessment Scales	Oil Price Uncertainty	Particle
Neurotrauma Effectiveness Research	Old Residential Buildings	Particle Swarm Optimisation
Neurotrophic Factor	Older-Adults	Particle Swarm Optimization
Neutrality	Omics	Particle-Laden Flows
Neutrinoless Double-Beta Decay	Omitted Variable Bias	Particle-Size Fractions
Neutrinos	Oncology	Particles
Neutron-Star	Online Assessment	Particulate
Neutrophil-Lymphocyte Ratio	Online Chatter Detection	Particulate Matter
New Framework	Online Flora	Pasc
New Taxa	Online Joint Estimation	Patch-Image Model
New Tools	Online Parameter Identification	Patent Knowledge
New Variant Strain	Online Research	Pathological Brain Detection
News	Oomycete	Pathological Complete Response
Nf-Kappa-B	Open Partial Nephrectomies	Pathology
Nicotinic Acetylcholine-Receptor	Open-Label	Pathways
Night Salivary Cortisol	Openness	Patient
Nighttime Light Data	Operational	Patient-Led Research
Nirsevimab	Operator	Pattern
Nitrogen	Opinion Mining	Patterns
Nitrogen Chemistry	Opportunities	Pavement Crack Detection
Nitrogen Deposition	Opportunity	Paxlovid (Tm)
Nitrogen Nutrition	Opposition -Based Learning	Pbbi Bearing
Nitrogen Reduction	Optical	Pca
Nivolumab	Optical Computing	Pd-1 Blockade
Nlr	Optical Design	Pd-L1
Nlrp3 Inflamasome	Optical Imaging	Pd-L1 Blockade
Nms	Optics	Pdes
Nn Training	Optimal Reference Tracking	Pea Protein Isolate
Nociceptors	Optimization	Peak
Noise	Optoelectronics	Pedestrian Detection
Noise Level	Oral Antiviral	Peer-To-Peer Computing
Noise Reduction	Oral-Bacteria	Pembrolizumab
Non-Darcy Medium	Orbital Coupled-Cluster	People
Non-Dominated Solutions Sorting Genetic	Ordinary And Partial Differential Equations	Peptide Identification
Non-Pharmaceutical Measures	Organ Failure	Perceived Fairness
Non-Renewable Energy Consumption	Organ Function	Perceived Impact
Nonalcoholic	Organic Amendments	Perceived Knowledge
Nonalcoholic Fatty Liver	Organic Pollutant Oxidation	Perception
Nonalcoholic Fatty Liver Disease	Organic Waste	Percutaneous Coronary Intervention
Noncommunicable Diseases	Organic-Carbon Stocks	Performance
Nonconvex Machine	Orthogonal Signal Correction	Performance Analysis
Nondestructive Estimation	Orthomcl	Period
Noninvasive Brain-Stimulation	Orthostatic Tachycardia Syndrome	Period-Luminosity
Nonlinear Dynamic Behaviors	Oscillation Spectroscopic Survey	Perivascular Spaces
Nonlinear Integro-Differential Model	Oscillatory Field Measurement	Permeability
Nonlinear Partial	Osmolytes	Permutation Test
Nonlinear System	Outbreak	Peroxidation
Nonlinear Time-Dependent Pdes	Outcome Measures	Personal-Care Products
Nonorthogonal Multiple-Access	Outcomes	Personality-Traits
Nonperturbative	Outlier Detection	Personalized
Nonrenewable	Output	Perspective
Nonrenewable Energy	Outsourced Agricultural Machinery	Pet
Nonsteroidal	Ovarian	Peucedanum
Novel Joint Transfer Network (Njtn)	Overall Survival	Pgpr Taxonomic And Functional Traits
Novelty Detection	Overdiagnosis	Pharmaceuticals
Nox Formation	Oxaliplatin	Pharmaceuticals And Personal Care Products
Npp-Viirs		

Pharmacokinetics	Post-Acute Sequelae Of	Propagation Losses
Pharmacotherapies	Post-Acute Sequelae Of Covid-19	Propensity Score
Phase-Change Materials	Post-Covid Syndrome	Proper
Phase-Change Memory	Post-Covid-19	Proper Motions
Phase-II	Post-Covid-19 Condition	Property
Phase-III	Post-Covid-19 Syndrome	Prophylaxis
Phased Assembly	Post-Marital Residence	Proposals
Phenol	Postmenopausal Patients	Prospective Effects
Phenotypic Characteristics	Postnatal-Growth	Prostate
Phenotypic Plasticity	Postoperative Prediction Model	Prostate Cancer
Phenotyping	Posttraumatic-Stress	Prostate-Cancer
Phosphorylation	Posttraumatic-Stress-Disorder	Protective Factor
Photocatalytic H 2 Production	Potassium Copper Hexacyanoferrate	Protein
Photon Imaging Camera	Potato	Protein Folding
Photons	Potent	Protein Identification
Photothermal Spectroscopy	Potential Transit Signals	Protein Language Model
Phylogenetic	Potential-Energy	Protein Secondary Structure
Phylogenetic Analysis	Potential-Energy Surfaces	Protein Stability
Phylogenetic-Relationships	Potentials	Protein Structures
Phylogenetics	Powder	Protein-Kinase-C
Phylogenomics	Powell-Eyring Fluid	Protein-Protein Interactions
Phylogeny	Power	Proteins
Physical Education	Power Systems	Proteomics
Physical Stability	Ppb-Level Detection	Protocol
Physical-Activity	Practice Parameter Update	Protocols
Physicians	Precision Medicine	Proton-Pump Inhibitor
Physics-Informed Deep Learning	Predict Axial Load	Provably-Secure
Physics-Informed Learning	Prediction	Provides
Physics-Informed Neural Networks	Predictions	Provides Insights
Phytochemicals	Predictive Control	Pso
Phytoplankton	Predictive Model	Psychiatric Comorbidity
Pi-Pi Interaction	Predictive Modelling	Psychiatric Illness
Picard'S Analysis	Predictive Models	Psychological
Pickering Emulsion	Predictive Soil Mapping	Psychological Impact
Pigeon	Predictors	Psychological-Factors
Pins	Prefrontal Cortex	Psychometric Properties
Planned Behavior	Pregnancy	Psychopathology Symptom Networks
Plant	Pregnant-Women	Psychosocial Support
Plant Disease Detection	Preoperative Chemoradiotherapy	Psychotherapies
Plant Diversity	Prescription	Public
Plant Essential Oils	Preserved Ejection Fraction	Public Acceptance
Plant Extracts	Pressure	Public Health
Plant Functional Traits	Pressure-Drop Characteristics	Public Health Interventions
Plant Interactions	Preterm Birth	Public Perceptions
Plant Photosynthetic Pathway	Prevalence	Public Subsidies
Plant Root Functions	Prevent Gastric-Cancer	Public Transportation
Plasma	Prevention	Public-Health
Plasma Science And Technology	Preventive Psychiatry	Public-Participation
Plastic Pollution	Prevents Bone Loss	Public-Private Partnerships
Plastic Products	Price	Publication Decisions
Plasticizers	Price Volatility	Pulmonary Arterial-Hypertension
Plastics	Primary Health Care	Pulsar
Plastome Evolution	Primary Prevention	Pulsars: General
Plate	Primary Therapy	Pulse Pressure
Platform	Primary Tumor	Pulse-Like Ground Motion
Platinum	Primary-Care	Purchase Intentions
Plr	Principal	Pure Torsion
Plus	Principal Component	Pv
Plus Chemotherapy	Principal-Element Alloys	Pyramid Network
Plus Lomustine	Principles	Pyrolysis Temperature
Pm2.5	Prior Knowledge	Pyrolysis-Gc
Pneumonia	Privacy	Qm
Point	Privacy Protection	Qualitative Research
Point Cloud Processing	Privacy-Preserved Data Sharing	Quality
Point-Of-Care	Proanthocyanidin Biosynthesis	Quality Assessment
Points	Probabilistic Bayesian Deep	Quality Characteristics
Poisson-Gamma	Probabilistic Sensitivity-Analysis	Quality Of Service
Polar Amplification	Probability	Quality-Of-Life
Polarization	Probe Wmap Observations	Quant Dx Assay
Political	Problematic Smartphone Use	Quantification
Political Communication	Procurement	Quantile Causality
Political Risk	Product	Quantile Regression
Political Trust	Product And Process Innovations	Quantile-On-Quantile
Pollution	Product Innovation	Quantitative-Analysis
Pollution Haven Hypothesis	Product Life-Cycle	Quantum
Pollution Reduction	Product Lifecycle	Quantum Dots
Polybrominated Diphenyl Ethers	Product-Service System (Pss)	Quantum Mechanics
Polycyclic Aromatic-Hydrocarbon	Productivity	Quarantine
Polymeric	Productivity Growth	Quasi-Natural Experiment
Polymers	Products	Question Answering
Polyploid	Professional Isolation	Questionable Research Practices
Polyps	Profile	Questionnaire
Population	Prognosis	Quiescent
Population Aging	Prognostic And Health Management	R Package
Population-Based	Prognostics	R-2
Population-Genetics	Prognostics Health Management	Race
Pornography-Use Disorder	Program	Race-Related
Porous Magnesium Scaffold	Programmed Necrosis	Racial Disparities
Porter Hypothesis	Programs	Racism
Portfolios	Progress	Radar Imaging
Pose Estimation	Progression	Radar Polarimetry
Positron-Emission-Tomography	Project	Radial Abundance Gradients
Post	Projections	Radial Direction
Post-Acute	Propagation	Radiation

Radical Nephrectomy	Renewable Energy	Rock
Radio Astronomy	Renewable Energy Consumption	Rock Bolts
Radio Frequency	Renewable Energy Investment	Rock Slide Blocking River
Radiomics	Renewable Energy Sources	Rollingwindow
Raman Microspectroscopy	Renewable Energy Technology Innovation	Rooftop Photovoltaic
Raman Spectroscopy	Renewable Energy Volatility	Rooftop Pv Generation
Random Forest	Renewable Energy-Consumption	Room-Temperature
Random Forest Regression	Rental Market	Root
Random Number Generator	Replication	Ross Ice Shelf
Random-Access Memory	Reporting Guideline	Rotating
Random-Field	Reporting Guidelines	Rothamsted Classical
Randomized	Repositories	Rotor
Randomized Controlled-Trials	Representation	Route Selection
Randomized Phase-Iii	Representation Learning	Rule
Randomized-Trial	Representations	Rules
Range Expansion	Reproducibility	Rxlr Motif
Ranking	Requiring Prolonged Observation	S Ets Pilot Policy
Rapid	Research	S-Adenosylmethionine
Rapid Detection	Research And Development	S-Measure
Rare	Research Culture	Saccharomyces-Cerevisiae
Rare-Earth-Elements	Research-And-Development	Safety
Rate Performance	Residency Training	Safety Vaccine
Rates	Residents	Salicylic-Acid
Rationalization Of Industrial	Residual Network	Salience
Reactive Force-Field	Residue Contacts	Saliency
Reactive Red-120	Resilience	Saliency Detection
Reactive Transport	Resistance	Salient
Reactor	Resistant Design	Salient Object
Real-Time	Resistant Prostate-Cancer	Salient Object Detection
Real-Time Detection	Resolution	Salp Swarm Algorithm
Real-Time Object Detection	Resolution Mass-Spectrometry	Sample
Real-Time Processing	Resource	Sample Set
Realized Volatility	Resource Curse	Sample Size
Receptive-Fields	Resource Management	Sample-Size
Receptor	Resource-Based	Sandy Soil
Receptor T-Cells	Resources Consumption	Sanjiang Plain
Receptor-Binding Domain	Respiratory	Sar
Recession	Respiratory Syncytial Virus	Sar Ship Detection Dataset (Ssdd)
Recognition	Respiratory Syndrome Coronavirus	Sars
Recombination	Respiratory Syndrome Outbreak	Sars Coronavirus
Recommendations	Respiratory-Syndrome-Coronavirus	Sars-Cov-2
Reconfiguration	Respondents	Sars-Cov-2 Infection
Reconstruction	Response Ratios	Sars-Cov-2 Neurotropism
Records	Response Surface Method	Satellite
Recovery	Response To Treatment	Satellite Data
Recruitment	Response-Surface	Satellite-Observations
Rectum Cancer	Response-Surface Methodology	Satellites
Recurrent	Responses	Saturation
Recurrent Neural Network (Rnn)	Resting-State Fmri	Saving Behavior
Recursive Feature	Restoration	Scalable Production
Recycled Glass	Restriction	Scale
Red Soil Region	Resuscitation	Scanning Time
Reduced Ejection Fraction	Retrofit	Scattering
Reduced Graphene Oxide	Return	Scattering Oriented Analysis
Reduced Order	Return On Assets	Scene Understanding
Reducing Carbon Emissions	Returns Evidence	Scenes
Reduction	Reveals	Scheme
Refuse-Derived Fuel	Reverberation Mapping Project	School Closure
Regenerative Braking	Review	School Closures
Region	Review-Systematic	School-Aged Children
Region Proposal	Rgb	Science
Regional Homogeneity	Rgb-D	Science Mapping
Regional Integration	Rgb-D Sod	Scientific
Registered Replication Report	Rhabdomyolysis	Scientific Progress
Registry	Rheumatic	Scoring Function
Regression	Rheumatoid	Scoring Functions
Regression Evaluation	Rheumatoid Arthritis	Screening
Regression Evaluation Rates	Rhythm	Sdof Systems
Regularity	Rice	Sdss-Iv
Regularization	Rice Farmers	Sdss-Iv Manga
Regulatory Network	Risk	Se Tibetan Plateau
Rehabilitation	Risk Estimation	Sea
Reinfection	Risk Factor	Sea Co2 Fluxes
Reinforced Concrete (Rc) Beams	Risk Factors	Sea-Ice Loss
Reinforced-Concrete Beams	Risk Perception	Sea-Surface Temperature
Reionization	Risk Prediction	Search
Reirradiation	Risk Premium	Seaweed
Relation Extraction	Risk-Adjustment	Seaweeds
Relative Income	Risk-Factors	Second Generation Panel
Relative Oscillator-Strengths	Risks	Secondary
Relativistic	River	Secondary Structure
Relaxation-Times	River Basin	Secreted Apolipoprotein-E
Reliability	River-Basin	Sector
Remaining Useful Life	Rivers	Sectorindices
Remaining Useful Life Prediction	Rna	Secular Trends
Remote Sensing	Rna Gene	Secure
Remote Sensing Image	Rna-Seq	Security
Remote Sensing Images	Rna-Seq Data	Sedentary Behavior
Remote Working	Road	Sedentary Behaviors
Removal	Roadmap	Sediment Yield
Renal Cell Carcinoma	Roads	Sediments
Renal-Cell	Robust Reversible	Seepage
Renal-Cell Carcinoma	Robust Watermarking	Segmentation
Renal-Transplant Recipients	Robustness	Seismic Data

Seismic Performance Selection	Slaughter	Spectrum Disorder
Selective Electrode	Sleep	Spectrum Disorders
Self -Centring Main Frame	Slime Mould Algorithm	Specular Reflection Learning Mechanism
Self-Powered Sensing	Slope Stability Prediction	Speed
Semi-Supervised Learning	Slow Feature Analysis	Sperm
Semaglutide 2.4 Mg	Small Effects	Spillovers
Semantic Change Detection	Small Firms	Spin-Waves
Semantic Segmentation	Small-Molecule	Spiral
Semantics	Smallpox Vaccination	Split Delivery
Semen Quality	Smart	Sponge-Derived Fungus
Semi-Leptonic Weak	Smart Cities	Spontaneous-Abortion
Semi-Supervised	Smart Contracts	Sport
Semi-Supervised Learning	Smart Devices	Spread
Semiarid Ecosystems	Smart Grids	Ssa-Elman
Semilinear Parabolic Equation	Smart Home	Stability
Sensing Image Fusion	Smart Textiles	Stacked Auto Encoder
Sensitivity	Smoke	Stacking
Sensor s	Smoking	Stage Manufacturing System
Sensor Drift	Smoothing Methods	Staging
Sensor Mapping	Smote	Stakeholder Management
Sensor Systems	Social Capital	Standard
Sensors	Social Defeat Stress	Staphylococcus-Aureus
Sentiment Analysis	Social Determinants	Star Formation
Sentinel-1A	Social Inequality	Star-Formation Histories
Separation And Reaction	Social Isolation	Star-Formation Rates
Sepsis	Social Media	Starch
Septic Shock	Social Network Analysis	Stars:
Sequelae	Social Network Large-Group Decision Making	State
Sequence	Social Networking	State Estimation
Sequence Database	Social Networks	State Of Health
Sequences	Social-Isolation	State-Of-The-Art
Serology	Sociodemographic	Statins
Seroprevalence	Socioeconomic	Statistical Evaluation
Serum 25-Hydroxyvitamin D	Socioeconomic Impact	Statistical Perspective
Serum Creatinine	Socioeconomic-Status	Statistical Power
Serum Uric-Acid	Soft Detector	Statistics
Server	Soft Tissue Grasping	Steam-Generation
Servers	Software	Stellar
Service	Soil	Stellar Mass Function
Service Intensity	Soil Carbon	Stem-Cell Population
Service User	Soil Loss	Stem-Cell Transplantation
Services	Soil Microbial Diversity	Stiffness Matrix
Set	Soil Organic-Carbon	Stimulation
Set Enrichment Analysis	Soil Physical	Stochastic
Sets	Soil Remediation	Stochastic Gradient
Severity	Soil-Landscape Model	Stochastic Gradient Descent
Sewage Sludge	Soil-Moisture	Stochastic Processes
Sex	Soil-Water	Stock Liquidity
Sex Reassignment Surgery	Solanum-Bulboscastanum	Stocking Rate
Sex-Chromosomes	Solar Energy	Strain
Sex-Differences	Solar Radiation Prediction	Strains
Sgt2 Inhibitors	Solar Still	Strategies
Sgt2-I	Solar-Cells	Strategy
Shap	Solid Modeling	Streamflow Forecasting
Shape	Solid Particles	Streaming Media
Shape Factor	Solid-Phase-Extraction	Street View
Shapes	Solubility	Street View Images
Shear-Strength	Sorption	Streetscape Greenery
Sheehan Disability Scale	Sorting	Strength
Sheep	Source	Strengthening Mechanism
Shell	South	Streptococcus-Pneumoniae
Shift	South Africa	Stress
Ship Pollutant Emissions	Southern-Ocean	Stress-Distribution
Ship-Borne Measurements	Soybean-Meal Diets	Stressful Life Events
Shop Rent	Sp-Nov	Stroke
Short-Term	Spain	Stroke Prevention
Shot	Sparse	Structural Basis
Shuffled Frog Leaping Algorithm	Sparse Autoencoder	Structural Changes
Siamese Network	Sparse Representation	Structural Equation Modeling (Sem)
Sigmoidoscopy	Spatial	Structural Equation Models
Signal Processing	Spatial Difference-In-Differences Model	Structural Performance
Sii	Spatial Durbin Model	Structural Racism
Silica-Gel	Spatial Presence	Structural Reliability
Similarity Analysis	Spatial Resolution	Structural Separation
Simpsons Paradox	Spatial Spillover Effect	Structural Variation
Simulation	Spatial Vegetation Patterns	Structural Walls
Simulation System	Spatial-Temporal Attention	Structure Measure
Sine Cosine Algorithm	Spatiotemporal	Students
Single	Spatiotemporal Evolution	Sub-Saharan Africa
Single-Atom Catalysts	Speaker Variation	Subjective Social-Status
Single-Cell	Speciation	Subjective Well-Being
Single-Cell Analysis	Species	Subsets
Single-Cell Genomics	Species Concepts	Subspace Identification
Single-Molecule Fret	Species Criteria	Substance Use Disorders
Singular	Species-Diversity	Substitution
Singular Orbit	Spectral Bias	Subtropical Reservoir
Singular Spectrum Analysis	Spectral Energy-Distributions	Subtypes
Sio2/H2O	Spectrometry Data	Success
Siri	Spectrometry-Based Metabolomics	Successive Refinement
Sites	Spectroscopic	Sulfate Radical-Advanced
Size	Spectroscopic Database	Summer
Skyrme Interaction	Spectroscopic Survey	Summer Maize
Slacks-Based Measure	Spectroscopy	Sunitinib
		Super-Earth

Super-Resolution	Text Classification	Trials
Superficial Siderosis	Texture Analysis	Triaxial Accelerometer
Supernova	Tfbsshape	Triboelectric Nanogenerator
Superresolution	Thalamic Activity	Triboelectric Nanogenerators
Supersonic Flows	The Plus Model	Triglyceride-Rich Lipoproteins
Supervised	Thematic Evolution	Triglycerides
Supervised Learning	Theory Of Planned Behavior	Triple -Negative Breast Cancer
Supply Chain	Theory Of Planned Behaviour	Tropical Forest Biomass
Support	Therapy	Tropical Pacific-Ocean
Support Vector	Thermal	Tropism
Support Vector Machine	Thermal Structure	Troponin
Surface	Thermal-Neutron-Capture	Tuberculosis
Surface Ocean Pco(2)	Thermochemical Conversion	Tumor Mutation Burden
Surface-Antigen	Thermochemistry	Tumor Progression
Surface-Area	Things	Tumor-Antigen
Surface-Plasmon Resonance	Three-Dimensional Displays	Tumor-Cells
Surgery	Three-Phase Co-Evolutionary Strategy	Tumor-Necrosis-Factor
Surgical	Three-Scroll Chaotic Attractor	Turbulence
Surgical Adjuvant Breast	Three-Scroll Lu-Like System	Turnover Frequency
Surrogate Model	Thrombectomy	Twitter
Surrogate Modeling	Thrombolysis	Two-Way Coupling
Surrogate Safety Measure	Ti3C2Tx	U Shaped
Surveillance	Ti3C2Tx Mxene	Ualcan
Survey	Tibetan Plateau	Uav Images
Surveys	Tight Control	UiO-66
Survival	Time	Uk Biobank
Survivors	Time Discretization	Ulcerative Colitis
Survivorship	Time Series	Ultra Filtration
Susceptibility	Time Synchronization	Ultra-High Risk
Susceptibility Loci	Time Trends	Ultralow Thermal-Conductivity
Sustainability	Time-Series	Uncertainty
Sustainable	Timetrees	Uncertainty Quantification
Sustainable Development	Tip Dating	Uneven Developments
Sustainable Green Finance	Titers	Unified Model
Swarm	Tm6Sf2	Union-Of-Pharmacology
Swarm Intelligence	To 8	Uniref
Swarm-Intelligence	To-Pay	Unit Roots
Switches	Tolerance	Unit-Root Tests
Symmetric Buildings	Tolerant	United-States
Symptomatology	Tomato	Unknown Planetary Boundary
Symptoms	Tool	Unobserved
Synaptic Plasticity	Toolkit	Unsupervised Anomaly Detection
Syndrome Coronavirus 2	Tools	Unsupervised Learning
Synthetic Aperture Radar	Top-Down Amplification	Update
Synthetic Aperture Radar (Sar)	Topology	Update Strategy
Synthetic Evaluation	Torsional Strength	Updated
System	Total Electron-Content	Upper Jinsha River
System Dynamics	Total Factor Productivity	Upscaling
Systematic Analysis	Total Knee Arthroplasty	Urban
Systematic Review	Total Knee Replacement	Urban Co2 Emissions
Systematic Reviews	Total Mesorectal Excision	Urban Expansion
Systematic-Errors	Total Parenteral-Nutrition	Urban Green Innovation (Ugi)
Systematics	Total-Energy Calculations	Urban Green Space
Systems	Trace	Urban Innovation
Systems Architecture	Tracking	Urban Urban Built-Up Area
T-Cell Immunity	Traction Motors	Urbanization
T-Cells	Trade	Us
T-Norms	Trade Openness	Usa
Tablets	Trading	Usage
Tabular Data	Training	Use Efficiency
Target	Training Data	Use Regression-Models
Target Detection	Trait-State Models	Useful Life Prediction
Task	Trajectory	User
Task Analysis	Transcription Factors	User Experience
Task-Force Report	Transcriptome	Uv Luminosity Function
Tau	Transcriptomics	Uv Luminosity Functions
Taxonomic Consistency	Transcripts	Vaccinate
Taxonomy	Transcultural Adaptation	Vaccination
Tdcs	Transfer Learning	Vaccination Intentions
Techniques	Transformation	Vaccine
Technological	Transformer	Vaccine Hesitancy
Technological Innovation	Transformers	Vaccine Uptake
Technological-Innovation	Transient	Vaccines
Technologies	Transition	Vaccinia Virus
Technology	Transition-Metal Carbides	Validation
Technology Spillover	Translation	Validity
Technology-Assessment	Transmembrane Domain Dimerization	Value-Added
Tecovirimat	Transmission	Value-At-Risk
Teenagers	Transparency	Variable
Telemedicine	Transplant Recipients	Variable Production Rate
Telework	Transplantation	Variable Selection
Temozolomide	Transport	Variables
Temperature	Transporter	Variant
Temperature Distributions	Travel	Variants
Temporal	Traveler	Variants Of Concern
Tension-Type Headache	Traveling Ionospheric Disturbances	Variational Inference
Terrestrial Ecosystem	Treatment	Vehicle Detection
Terrestrial-Based	Treatment Patterns	Vehicle Routing Problem
Tert Promoter Mutations	Treatment-Resistant Depression	Vehicle-To-Grid
Test Errors	Tree-Based Models	Velocity Dispersion Profiles
Test-Negative Design	Trees	Venous Thromboembolism
Testing	Trend	Verification
Tetracycline	Trends	Version Space Learning
Text	Trial	Versus-Host-Disease

Vibrational Properties	Waste-To-Energy	Winter Garlic
Vibrations	Waste-Water	Winter Wheat
Vicarious Racism	Waste-Water Treatment	Winter-Wheat
Video Superresolution	Wastewater	Wireless Communication
Videos	Water	Women
View	Water Desalination	Women 45 Years
Vigilance	Water Erosion	Women'S Empowerment
Vineyards	Water Quality	Womens Autonomy
Viral Pneumonia	Water Treatment	Wood-Rotting Fungi
Viral-Hepatitis	Water Treatment-Plant	Woody Plant Encroachment
Virtual	Water-Adsorption	Work
Virtual Collection	Water-Leaving Radiance	Workers
Virtual Machines	Water-Stress	World-Health-Organization
Virtual Reality	Watermelon	Worldwide
Virtual Worlds	Wave	Wsn
Virus	Wave Energy	Wuhan
Virus Ah1N1/09	Waveform	Wwtps
Viruses	Wavelet Packet	Wx Gene
Visible Light Communication	Wavelet Transform	X-Ray Images
Vision	Waxy Gene	X-Ray-Absorption
Vision Transformer	Wearable Device	X-Ray-Emission
Visual-Attention	Web	X-Rays: General
Visualization	Web Server	Xpert Hiv-1
Vitamin D	Web-Based Tool	Xps
Vitamin-D Deficiency	Website	Yangtze-River Delta
Vitamin-D Supplementation	Weight-Loss	Yellow River
Vitamins	Whale Optimization	Yh3
Viticulture	Whale Optimization Algorithm	Yield
Vivo Lung Perfusion	Wheat	Yield Prediction
Vocs Adsorption	White	Yolov3
Volatile	Who 2021 Global Tb Report	Yolov5
Volatility	Wide Association	Young Adults
Volatility Forecasting	Wide-Field Survey	Young People
Volcanic Eruption	Wild	Young Star-Clusters
Voltage-Gated	Wild-Land Fires	Young-Adults
Vulnerability	Wild-Type	Zero Valent Iron
Wage	Wildfire Detection	Zero-Sum Games
Waist	Willingness To Pay	Zeta-Valence Quality
Waste	Wind Erosion	Zinc -Ion Battery
Waste Biomass	Wind Power Forecasting	Zinc-Oxide Nanoparticles
Waste Glass	Wind Speed Forecasting	Zymoseptoria-Tritici
Waste Management	Winter Canola	