International Journal of Big Data Mining for Global Warming
Vol. 3, No. 1 (2021) 2150005 (15 pages)

DOI: 10.1142/S2630534821500054

DATA ANALYTICS FOR CLIMATE AND ATMOSPHERIC SCIENCE

STAVROS DEMERTZIS[*,§], VASILIKI DEMERTZI[†,¶] and KONSTANTINOS DEMERTZIS[‡,||]

[*]*School of Spatial Planning and Development*
Faculty of Engineering, Aristotle University of Thessaloniki
54124 Thessaloniki, Greece

[†]*Computer Science Department, School of Science*
International Hellenic University
University Campus, 65404 Kavala, Greece

[‡]*Department of Physics, Faculty of Sciences*
International Hellenic University
St. Lucas Campus, 65404 Kavala, Greece
[§]*demertzs@plandevel.auth.gr*
[¶]*vademer@teiemt.gr*
[||]*kdemertzis@teiemt.gr*

Received 24 February 2021
Revised 9 March 2021
Accepted 1 May 2021
Published

Global climate change has already had observable effects on the environment. Glaciers have shrunk, ice on rivers and lakes is breaking up earlier, plant and animal ranges have shifted and trees are flowering sooner. Under these conditions, air pollution is likely to reach levels that create undesirable living conditions. Anthropogenic activities, such as industry, release large amounts of greenhouse gases into the atmosphere, increasing the atmospheric concentrations of these gases, thus significantly enhancing the greenhouse effect, which has the effect of increasing air heat and thus the speedup of climate change. The use of sophisticated data analysis methods to identify the causes of extreme pollutant values, the correlation of these values with the general climatic conditions and the general malfunctions that can be caused by prolonged air pollution can give a clear picture of current and future climate change. This paper presents a thorough study of preprocessing steps of data analytics and the appropriate big data architectures that are appropriate for the research study of Climate Change and Atmospheric Science.

Keywords: Data analytics; Machine Learning; climate change; atmospheric science; air pollution; air quality.

1. Introduction

Climate change is thought to be responsible for the frequent occurrence of droughts, floods and the rise of infectious diseases such as malaria. It is also often attributed to

[||]Corresponding author.

extreme or dangerous weather events, and among its long-term effects is rising sea levels, with damage to crops and infrastructure, which can have incalculable socio-economic consequences. In particular, a wide range of effects of climate change are found in agriculture, fisheries, soil, biodiversity, water resources, coastal zones, heat- and cold-related mortality, flood damage and widespread forest fires (Barrie Pittock, 2009).

Extreme weather events have also changed in frequency and intensity. Heat strokes have become more common in most land areas and the frequency of heavy weather precipitation has increased in most areas (Law, 2012). The atmosphere and the ocean are volatile components, which interact with each other and significantly affect the climate system. High temperatures and increased water vapor levels create the right conditions for the occurrence of extreme phenomena (O'Brien, 2014).

In particular, with the increase of the human population and the production process over the years, the negative effects on the immediate environment have increased, which is often responsible for the harmful effects on the health of living organisms and ecosystems (Barrie Pittock, 2009). One of the most common forms of environmental pollution is air pollution. Air pollution is characterized by the presence of air pollutants in quantity, concentration or duration, which can cause alteration of the structure, composition and characteristics of atmospheric air.

The main sources of air pollution are associated with anthropogenic activities, are located mainly in urban areas and are related to energy production, transport, industry, but also heating of buildings, technical projects and households (Ziska and Dukes, 2014). The pollution is mainly caused by pollutants such as nitrogen oxides caused by the photochemical cloud, sulfur which reacts with cloud water vapor to produce acid rain, carbon which contributes to the greenhouse effect and soot produced by incomplete combustion such as non-combustible hydrocarbons and their admixture with gases (Rahel and Olden, 2008).

Real-time monitoring of fluctuations and, above all, the prediction of levels of air pollution (Seinfeld and Pandis, 1998), as well as the search for relationships between pollutants and other parameters that affect them, such as their correlation with other primary and secondary pollutants or meteorological factors, which is based on sophisticated decision-making methods, form one of the most important topics in environmental science and research (McClanahan and Cinner, 2010).

2. Literature Review of Big Data from Climate and Atmospheric Science

A link between big data and climate change has long been noted (Huisingh and Zhang, 2019). Machine Learning coupled with advanced earth observation is one of the ways in which this is being implemented (Demertzis *et al.*, 2020). Big Earth observation data have played an important role in fields such as quantifying the global forest changes (Hansen *et al.*, 2014), tracking global urban expansion (Liu *et al.*, 2019) and mapping land use and land cover (Felipe-Lucia *et al.*, 2020). These data are needed to produce novel data-stewardship reference models (Albani and Maggio, 2020), efficient data indexing and retrieval methods (Qu *et al.*, 2020), effective methods of improving data quality (Wang *et al.*, 2020) and data mining (Dumitru *et al.*, 2020) and high-performance data processing systems (Zhou *et al.*, 2020).

On the other hand, Machine and Deep Learning techniques are the most effective methods for analyzing big data from environmental science (Demertzis and Iliadis, 2020). Specifically, Deep Learning methodologies have significantly contributed to the evolution and development of hyperspectral and panchromatic images analysis and classification (Petersson *et al.*, 2016). Chen *et al.* (2014) use a hybrid method which combines stacked Autoencoder, Principal Component Analysis (PCA) and Logistic Regression to perform hyperspectral data classification. Tao *et al.* (2015) use a sparse-stacked Autoencoder to efficiently represent features from unmarked spatial data, and then learnt features are fed into a linear SVM for hyperspectral data classification. Various 1D (Kussul *et al.*, 2017) and 2D (Makantasis *et al.*, 2015) CNN architectures have been proposed from time to time to encode spectral and spatial information.

The latest and most sophisticated proposal concerns 3D CNN (Chen *et al.*, 2016) in which the third dimension refers to the time axis resulting in a spatio-temporal architecture in the spectral classification. In 3D CNNs, the convolution functions are spatial–spectral, whereas in 2D CNNs, they are spatial only. Compared to 1D and 2D CNNs, 3D CNNs can obtain better spectral information thanks to 3D convolution functions.

In addition, Mou *et al.* (2018) have proposed a novel network architecture, which is a fully Convolutional/Deconvolutional network, for unsupervised spectral–spatial feature learning of hyperspectral images, which is able to be trained in an end-to-end manner. Specifically, the proposed architecture is based on the so-called encoder–decoder paradigm. The input 3D hyperspectral patch is first transformed into a typically lower dimensional space via a convolutional subnetwork (encoder). Then it is expanded to reproduce the initial data by a de-convolutional subnetwork (decoder).

3. Data Analytics

Beyond their management, the biggest modern challenge for large-scale data is to analyze it in a functional way, so that the hidden knowledge contained in the information can be extracted. For example, using pattern recognition methods, it is possible to identify market trends or patterns, identify unknown correlations, customer preferences and other useful business information to achieve behavioral prediction, in order to make optimal decisions (Valafar, 2002).

The current trends in data analytics, in which it should be noted that in each case of methods for approaching the future forecasts, their degree of difficulty and their valuation increase exponentially, concern descriptive analytics, regulatory analytics, diagnostic analytics and predictive analytics.

In conclusion, large-scale data analysis techniques, especially the real-time analysis techniques, have significant effects on improving product flexibility and quality, energy efficiency and improved equipment maintenance through preventive measures and maximizing benefits in the whole cycle of the production process.

4. Architectures

In the case of classical computing architectures, data is treated as static (batch data), assuming that all data is available whenever needed, such as during the training of a learning algorithm (Bengio, 2009). In these architectures a final model is produced, using all available data which are examined simultaneously.

In contrast to the cases of knowledge extraction and application of models for optimal decision-making in real-time (Demertzis *et al.*, 2017a), architectural implementations using algorithms are required, such as Machine Learning algorithms, which learn from a series of data provided over time, while having the ability to be constantly updated as each new data stream (data stream processing) arrives (Iliadis *et al.*, 2018).

Naturally and easily understood, some of the most important problems that arise during the extraction of knowledge from data streams are related to the high speed at which information flows and the natural tendency of data to evolve from time to time thus resulting in the degradation of classifiers due to the constant change of information (concept drift) (Demertzis *et al.*, 2017b). Flow analysis algorithms are also challenged and controlled by a variety of possible factors related to both their reliability and accuracy of categorization, as well as the memory consumption, resource constraints and legitimate processing times they can meet.

The management and analysis needs of a large volume of static and data streams, combined with real-time processing requirements, laid the foundation for architectural standardization of big data analysis systems (Anezakis *et al.*, 2017).

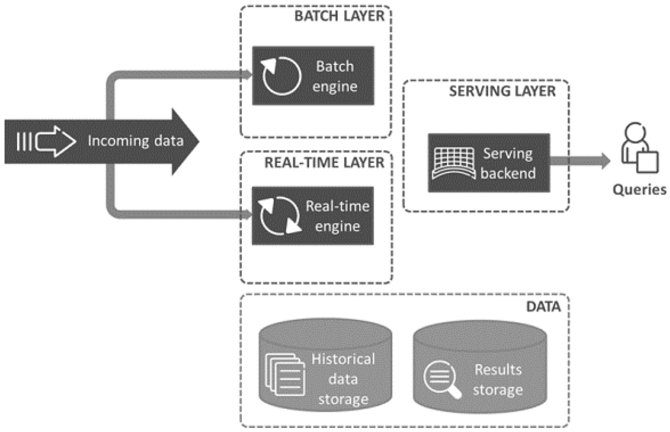
The two most important architectures that can meet modern dataset analysis requirements are Lambda and Kappa.

4.1. *Lambda*

Lambda (Demertzis *et al.*, 2019) is a data processing architecture designed to handle bulk data using batch data and data streams. This approach seeks to balance latency, throughput and fault-tolerance using the batch processing method to provide complete and accurate views of historical data while using real-time data stream processing to provide views of incoming data. The two projection outputs can be connected before the final presentation of the data or the final decision.

A graphic representation of the Lambda architecture is presented in Fig. 1.

The timeliness of the Lambda architecture is related as mentioned above to the development of big data, the analysis of this data in real-time and the attempt to mitigate the erroneous evaluation cases. It is also important to say that this architecture



Source: https://ww.ericsson.com/en/blog.

Figure 1. Lambda architecture.

is intended for receiving and processing data with time markings that are attached and function in addition to the existing data, while in no case seek to replace this data.

4.2. *Kappa*

Respectively, the continuous progress of technology and the continuous upgrade of the systems that are integrated into the infrastructures propose the production of data inflows, a fact that imposes their processing and analysis exclusively in real-time. As a result of the above formulation, there are requirements for serious technologies as well as specialized procedures and utilization of data flows, in order to predict the behavior and make optimal decisions in environments of dynamic displacement and feedback.

One of the most important and well-known architectures that have proven to be suitable for many of the challenges associated with analyzing large datasets and revealing hidden knowledge from data streams is the so-called Kappa architecture (Demertzis *et al.*, 2018). In this architecture, there is a real-time engine that undertakes the analysis of one data stream at a time. For each new sample, a small gradual update of the model takes place, which gradually improves as it receives more data streams.

Data streams are endless data that are produced by multiple network infrastructures, such as sensors, IoT equipment, etc. A typical example for streaming data is the data that can be collected from air pollution monitoring queries at high traffic rates. Streaming data should be handled in sequence and incrementally or over sliding time windows. Due to their reliance on strict time constraints and their more general availability, they are selected for detailed and specialized data processing techniques that can lead to multiple levels of revelation of the hidden knowledge they may contain. In addition, these data need to be processed without accessing all the rest of the data. In addition, it should be considered that concept drift may occur in the data, which means that the stream properties may transform and alternate over time. It is frequently used in big data projects, in which data streams are quickly produced by several dissimilar sources.

Streaming data and data generated by dynamic environments have led to some of the most robust research areas of the new era. Stream processing techniques are used by Machine Learning applications on data streams for real-time analysis and knowledge extraction under displacement and feedback environments. In the case of stream processing techniques by Machine Learning, the algorithms are controlled by a variety of possible shifting modes and constraints related to memory consumption, resource limitation and processing time. In this category, the available data are scaled in a sequential order and used for training and forecasting by calculating the error in each iteration. The aim of the algorithms in this category is to minimize the cumulative error for all iterations. We consider that the intention of supervised learning using the square loss function is to minimize the empirical error calculated by the following function:

$$I_n[w] = \sum_{j=1}^{n} V(\langle w, x_j \rangle, y_j) = \sum_{j=1}^{n} (x_j^T w - y_j)^2, \tag{1}$$

where $x_j \in R^d, w \in R^d$ and $y_j \in R$. Let there be a data table of dimensions $Xj \times d$ and a target values table of dimensions $Yj \times 1$ as they are defined after the entrance of the first i data points.

Let us suppose that the covariance table $\Sigma_i = X^T X$ is reversible, and $f^*(x) = \langle w^*, x \rangle$ is the ideal result for the linear least squares problem, as shown in the following equation:

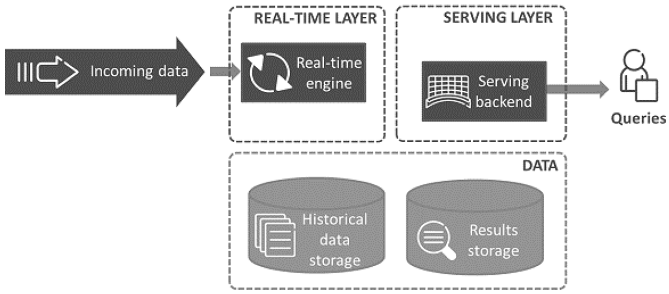
$$w^* = (X^T X)^{-1} X^T \Upsilon = \Sigma_i^{-1} \sum_{j=1}^{i} x_j y_j. \tag{2}$$

The calculation of the table $\Sigma_i = \sum_{j=1}^{i} x_j x_j^T$ has a time complexity of $O(id^2)$. Reversing the $d \times d$ table has a time complexity of $O(d^3)$, whereas the rest of the multiplication requires a time complexity of $O(d^2)$, producing an overall complexity of $O(id^2 + d^3)$. If we consider that n is the set of points in the dataset, $i = 1, 2, \ldots, n$, and it is essential to re-calculate the result after the arrival of each new data vector, we obtain a total time complexity of $O(n^2 d^2 + nd^3)$.

It is important here to mention that a Machine Learning stream processing is appropriate in cases where it is required to dynamically adapt the procedure to new standards or data, or when the streams are produced as a function of time, as in the case of the research of the climate analytics.

Essentially, the Kappa architecture is based on online or incremental learning algorithms, to solve real-time data analysis problems. A graphic representation of this architecture is presented in Fig. 2.

A key advantage of this architecture is its requirements for significantly smaller data storage space and system resources in general. Reinforcing the minimum resource requirement means the fact that by updating the model using only the latest data, the



Source: https://www.ericsson.com/en/blog.

Figure 2. Kappa architecture.

system does not need to store a large amount of data in memory and therefore significantly reduces its requirements for hardware infrastructure. This architecture also allows fast model updates free of computational complexity, so it is considered the most suitable option for applications with strict time budgets, such as real-time audio/video processing or applications where the response time should be in the order of milliseconds.

Finally, the flexible layout of this architecture adapts better to the data changes involved in a noisy field, which escalates over time. In particular, many of the learning methods used to analyze data streams have a "memory erase" property that allows the setting of a framework, usually a timeline, on the basis of which learning methods "forget" the past. In this way, the gradual attenuation of the importance of past data is achieved, a fact that allows the model to automatically adapt to the dynamic changes of its environment. This feature allows the model to operate safely even in the event of sudden changes that may indicate trends or even noise in the input data, e.g., extreme values.

Respectively, however, this feature is associated with a serious disadvantage of flow analysis methods called "Catastrophic Interference" or "Catastrophic Forgetting". It concerns the tendency of an algorithm to completely forget the information it had learnt during its training process, in the case of mnemonic connection models.

5. Data Processing

Data analysis is the process of inspecting, refining, transforming and modeling data with the aim of discovering hidden knowledge, drawing conclusions and making optimal decisions. It has multiple aspects and approaches, including a variety of techniques for applying them in different areas of business and science. It is commonly referred to as the process of collecting raw data, analyzing it in order to answer questions, test hypotheses, or refute theories, and process them properly to turn them into information useful for decision-making (Bishop, 2004).

Data analysis also involves the checks and preparation work that need to be done to get rid of the original data from various kinds of problems such as conflicting information, coding inconsistencies, field naming and units of measurement, and more important problems such as the existence of lost values, noise and extreme values, but also in dealing with special requirements that need the transformation of data, such as the discretization, normalization, reduction of their dimensions or the selection of the most appropriate characteristics. This process is particularly important because if the quality of the data is not ensured, quality and accurate decisions will not be made or untrue results will be produced.

In general, the data analysis process includes the following distinct steps, which in any case are necessary and should concern the analyst, but depending on the quality of the dataset, he/she will decide which of the following actions should proceed (Chizi and Maimon, 2005; Deng and Yu, 2014; Deselaers *et al.*, 2009).

5.1. *Data collection*

Data is usually the result of research or collection from sensors in the environment, such as traffic cameras, satellites, recorders, etc. It can also be the result of recording knowledge in a field, data collected through interviews, downloads from online sources, etc.

5.2. *Missing values*

Unavailable information is usually due to a variety of factors such as incorrect entry or inadvertent deletion of values by a registrar, hardware or software failure. They are one of the most common and serious data problems encountered in real applications, as they are very likely to disorient algorithms and lead to untrue results. Some of the possible ways to deal with lost values are to search and enter the actual value, delete the entire line, replace the lost values with a fixed value (e.g., null value), replace the lost values with the middle or most common value, replace the lost values with every possible value, replace the lost values by a reserved probability distribution and predict the lost values.

However, it is important to note that all methods of replacing lost values can cause discrepancies in the data, as the replacement price is probably not correct, but in most cases, there is no alternative.

5.3. *Outliers or extreme values (noise)*

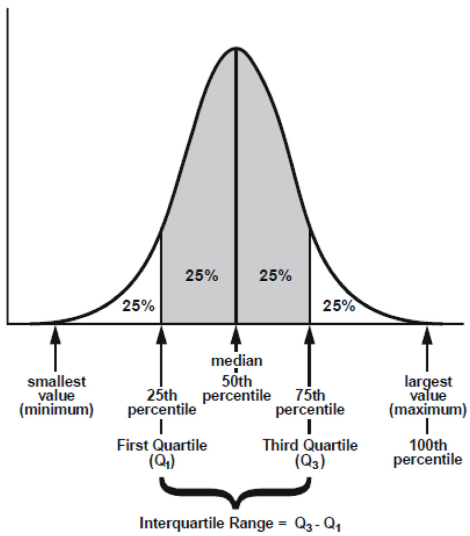
Noise refers to unwanted distortions in datasets, either systematic or completely random. Since data are usually expressed as sequences of numbers, noise is nothing more than numerical values that are added to and subtracted from the data. Noisy are the data that contain incorrect values and value-exceptions, i.e., values that do not provide useful information in the analysis. There are two tactics for dealing with noise. The first is based on replacing all numeric values with other appropriate values. The second is based on locating outliers or extreme values.

An extreme value is defined as a point that is very far from the mean value of the corresponding random variable representing a feature. Samples with characteristic values very different from the mean value produce significant errors, especially if they are the result of noise during the measurement process, which can cause catastrophic results. Distance is measured relative to a threshold, which is usually a multiple of the standard deviation. For a random variable following a normal distribution, a positive or negative distance equal to three times the standard deviation covers 95% of the points, and a distance equal to three times the standard deviation covers 99% of the points.

If the number of extreme values is small then either the values remain and are modified or these samples are simply discarded, which is the most popular tactic. One of the most popular methods of finding extreme values is performed with the technique of the Interquartile Range (IQR; see Fig. 3).

5.4. *Normalization*

Normalization is the process of transforming data, in which numeric values are replaced by the corresponding ones, but which are in a certain range of values. This process is usually performed to address problems related to the operation or performance of the algorithms. For example, some algorithms perform better when the input values fluctuate in the range of [0,1], while in algorithms that calculate the distances between observations, the normalization of values is required, as a problem is observed that variables with large values essentially determine the distance of the observations, while the variables with small values affect the distance little and finally, play no role in the calculation of the result.



Source: http://makemeanalyst.com/.

Figure 3. Interquartile Range.

There are several methods for normalizing numerical values, with the most commonly used being minimum–maximum normalization and decimal-scale normalization.

5.5. *Feature reduction*

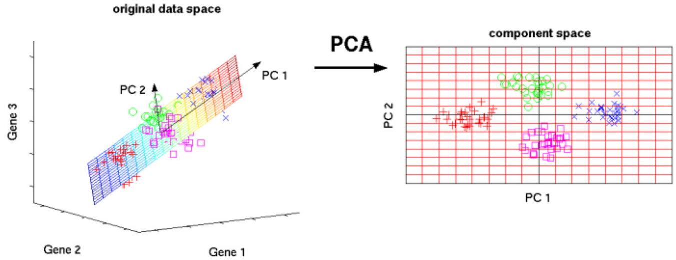
In most cases, a set of data can contain too many features, which may be related to each other, provide irrelevant information to the problem or produce noise, which reduces the quality of the data. Also, if the vector attribute space has many dimensions (i.e., many attributes), the volume of this space increases very fast, so the data for the problem will be sparse, creating problems in the methods that try to achieve statistical significance. The amount of data needed to be considered dense increases exponentially in relation to the dimension of the feature space. This phenomenon is also known as "the curse of dimensionality".

It should also be noted that a large number of features increases the number of parameters of the learning system, and therefore its complexity, without this meaning that it will have a correspondingly better performance. Because of these considerations, the number of features should be kept as small as possible to achieve high system performance. The solution to these problems is provided by dimensional techniques, which offer an efficient solution to managing multidimensional data, as they look for a low-dimensional structure in multidimensional data.

Preprocessing is considered necessary in such cases, as the distances between the data in the reduced space are calculated faster, the size of the dataset is reduced, the data structure is revealed which remains hidden in the original multidimensional space and the efficiency of the algorithms is significantly improved. The most well-known linear dimensional technique is the Principal Component Analysis, see Fig. 4.

5.6. *Feature selection*

This is the process of optimally selecting a subset of existing features without transformation, in order to maintain the most important of them, with the objective of reducing their number, and at the same time to retain as much useful information as



Source: http://www.nlpca.org/.

Figure 4. Principal Component Analysis.

possible. This step is crucial because if features with low resolution are selected, the resulting learning system will not perform satisfactorily, while if features that provide useful information are selected, the system will be simple and efficient. One of the strategies that can be followed is to examine the characteristics one by one through a measure of class reparability and to reject those that have a small resolution. The aim is to select these characteristics that lead to long distances between groups of samples and small variations between the same group.

This means that attributes should get distant values for different classes and close values for the same class, a strategy known as filtering. One of the most popular filtering methods is Forward Selection. Also, another approach to feature selection is achieved by examining the various combinations of features available and controlling those combinations that lead to higher performance, regardless of the quality of the individual features, an approach called Wrapping.

5.7. *Resampling*

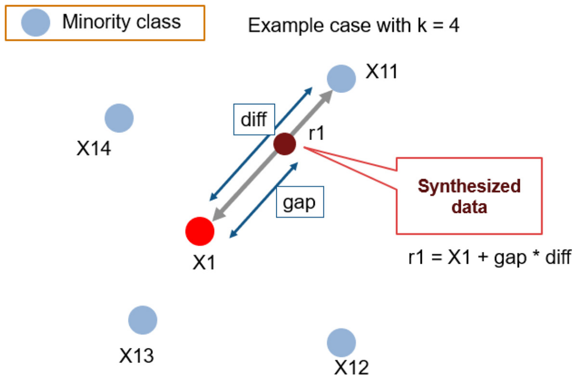
There are many cases in which a dataset is considered imbalanced, as the classes in a classification problem are not evenly distributed, i.e., there are many more snapshots of one class than the others. The problem that arises is that unbalanced sets of data tend to produce high prediction accuracy in the class with the majority of snapshots and correspondingly very low accuracy for the class with the fewest snapshots. This is because the classification algorithms used to assume that classes are evenly distributed across the data and aim to minimize the overall classification error to which the class with the fewest snapshots contributes little.

In addition, they assume that errors from different classes have the same cost. Random repetition of existing snapshots can increase the likelihood of the classifier being overfitting into very specific examples, as exactly the same copies are created and the rules generated by the classifier may be accurate, but in fact, only apply to a reproduced example (generalization is not achieved). At the data level, the various solutions proposed are based on the Oversampling sampling process. According to this method, the snapshots of the minority class are increased in order to achieve balance.

The process can be accomplished by randomly reproducing the existing snapshots of the minority class. But in this case, the decision-making surface creates conditions of consistency for the whole education, which does not produce accurate predictions in new data. Unlike the Synthetic Minority Oversampling Technique — SMOTE (Fig. 5), artificial samples are created, performing specific processes on the data of the minority class.

5.8. *Exploratory data analysis*

This process usually concerns statistical analysis procedures of the dataset, in order to understand the data, but also their variation in relation to the observations they describe, e.g., the average or median may be generated to help understand the data.



Source: https://www.analyticsvidhya.com.

Figure 5. Synthetic Minority Oversampling Technique.

Also, data visualization is the graphical representation of the data that can explain some phenomena in the collection or fluctuation of the observations, offering essentially the revelation of the hidden knowledge that the data may contain.

On the other hand, mathematical modeling and analysis of elements of the set can also reveal some hidden knowledge or some pattern or some relationship between observations, which can contribute to the actual analysis of the set; for example, correlation or causation analysis or even the finding of a characteristic error. There is some residual error depending on model accuracy (i.e., Data = Model + Error). Finally, data product concerns the process of application of algorithms, e.g., Machine Learning, for the prediction or approximation of a function based on the recorded data and the implementation of a model capable of providing knowledge in the field of research.

6. Conclusion

The basic components of climate change, which compose a multidimensional and extremely complex modern environmental problem, have become a global threat, with its effects becoming more and more visible in the daily life of ordinary people. The conditions that amplify climate change, but also its inherent effects, are very serious environmental problems of the highest priority and as it is understood, the timely, valid and with advanced technological systems tackling this phenomenon, the conditions and phenomena that enhance it but also of its effects, is a vital process. This can slow down the uncontrolled expansion of the problem, increase the likelihood of eliminating its difficulties and significantly enhance environmental safety, avoiding the need for costly and long-term remediation efforts.

Based on the specific need, specialized solutions and systems are required, which should be equipped with new, powerful and efficient analytics algorithms, in order to be able, through a flexible and easy-to-use operating environment, to complete specialized processes for modeling of complex environmental problems related to climate change. Using data science techniques, such as Machine Learning, we can build reactive, responsive decision-support systems to exploit incoming data on climate status, exploiting situational intelligence.

Also, the systems in question should be able to give accurate representations of the behavior of the actors involved in an environmental-scale simulation, to perform specialized techniques for analyzing the hidden knowledge that lies behind the data, and to be able to predict situations that can cause serious difficulties. Emerging technologies and scientific advances are moving at such a pace that architects, civil engineers, designers and developers are racing to deliver their benefits into our built environment. We can map natural capital and demonstrate how to preserve and enhance it. For example, diverse environmental analytics skills and an innovative product suite can allow quick simulation and assessment of energy scenarios supporting environmental strategic. All of the above actions can effectively support the optimal decision-making related to the problem of climate change.

In this paper, we have discussed the use of an innovative method for analysis and modeling of Climate and Atmospheric Science. More specifically, we analyze the necessary data analysis process of inspecting, refining, transforming and modeling data with the aim of discovering hidden knowledge, drawing conclusions and making optimal decisions. These steps might help to produce novel data-driven frameworks for real-time biodiversity monitoring, early detection of harmful invasive species, long-term environmental control systems, consequences containment measures and effective application of the international conventions of the climate change.

References

Albani, M and I Maggio (2020). Long time data series and data stewardship reference model. *Big Earth Data*, 4(4), 353–366. doi: 10.1080/20964471.2020.1800893.

Anezakis, V-D, L Iliadis, K Demertzis and G Mallinis (2017). Hybrid soft computing analytics of cardiorespiratory morbidity and mortality risk due to air pollution. In *ISCRAM-med 2017: Information Systems for Crisis Response and Management in Mediterranean Countries*, I Dokas, N Bellamine-Ben Saoud, J Dugdale and P Díaz (eds.), Lecture Notes in Business Information Processing, Vol. 301, pp. 87–105. Cham: Springer. doi: 10.1007/978-3-319-67633-3_8.

Barrie Pittock, A (2009). *Climate Change: The Science, Impacts and Solutions*, 2nd edn. Clayton, Australia: CSIRO Publishing.

Bengio, Y (2009). Learning deep architectures for AI. *Foundations and Trends in Machine Learning*, 2(1), 1–127.

Bishop, CM (2004). *Neural Networks for Pattern Recognition*. New York, NY: Oxford University Press.

Chen, Y, Z Lin, X Zhao, G Wang and Y Gu (2014). Deep learning-based classification of hyperspectral data. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 7(6), 2094–2107.

Chen, Y, H Jiang, C Li, X Jia and P Ghamisi (2016). Deep feature extraction and classification of hyperspectral images based on CNN. *IEEE Transactions on Geoscience and Remote Sensing*, 54(10), 6232–6251.

Chizi, B and O Maimon (2005). Dimensions reduction and feature selection. In *Data Mining and Knowledge Discovery Handbook: A Complete Guide for Practitioners and Researchers*, pp. 1189–1201. Boston: Springer.

Demertzis, K and L Iliadis (2020). GeoAI: A model-agnostic meta-ensemble zero-shot learning method for hyperspectral image analysis and classification. *Algorithms*, 13(3), 61. doi: 10.3390/a13030061.

Demertzis, K, L Iliadis and V-D Anezakis (2017a). A deep spiking machine-hearing system for the case of invasive fish species. In *Proceedings of the 2017 IEEE International Conference on INnovations in Intelligent SysTems and Applications (INISTA)*, pp. 23–28. Piscataway: IEEE. doi: 10.1109/INISTA.2017.8001126.

Demertzis, K, L Iliadis and V-D Anezakis (2017b). Commentary: *Aedes albopictus* and *Aedes japonicus* — two invasive mosquito species with different temperature niches in Europe. *Frontiers in Environmental Science*, 5, 85. doi: 10.3389/fenvs.2017.00085.

Demertzis, K, L Iliadis and V-D Anezakis (2018). A dynamic ensemble learning framework for data stream analysis and real-time threat detection. In *ICANN 2018: Artificial Neural Networks and Machine Learning*, V Kůrková, Y Manolopoulos, B Hammer, L Iliadis and I Maglogiannis (eds.), Lecture Notes in Computer Science, Vol. 11139, pp. 669–681. Cham: Springer. doi: 10.1007/978-3-030-01418-6_66.

Demertzis, K, L Iliadis and V-D Anezakis (2019). A machine hearing framework for real-time streaming analytics using Lambda architecture. In *EANN 2019: Engineering Applications of Neural Networks*, J Macintyre, L Iliadis, I Maglogiannis and C Jayne (eds.), Communications in Computer and Information Science, Vol. 1000, pp. 246–261. Cham: Springer. doi: 10.1007/978-3-030-20257-6_21.

Demertzis, K, L Iliadis and E Pimenidis (2020). Large-scale geospatial data analysis: Geographic object-based scene classification in remote sensing images by GIS and deep residual learning. In *EANN 2020: Proceedings of the 21st EANN (Engineering Applications of Neural Networks) 2020 Conference*, L Iliadis, P Angelov, C Jayne and E Pimenidis (eds.), Proceedings of the International Neural Networks Society, Vol. 2, pp. 274–291. Cham: Springer.

Deng, L and D Yu (2014). Deep learning: Methods and applications. *Foundations and Trends in Signal Processing*, 7(3–4), 197–387.

Deselaers, T, S Hasan, O Bender and H Ney (2009). A deep learning approach to machine transliteration. In *Proceedings of the Fourth Workshop on Statistical Machine Translation*, pp. 233–241. New York, NY: ACM.

Dumitru, CO, G Schwarz, A Pulak-Siwiec, B Kulawik, M Albughdadi, J Lorenzo and M Datcu (2020). Understanding satellite images: A data mining module for Sentinel images. *Big Earth Data*, 4(4), 367–408. doi: 10.1080/20964471.2020.1820168.

Felipe-Lucia, MR *et al.* (2020). Land-use intensity alters networks between biodiversity, ecosystem functions, and services. *Proceedings of the National Academy of Sciences of the United States of America*, 117(45), 28140–28149.

Hansen, M, P Potapov, B Margono, S Stehman, S Turubanova and A Tyukavina (2014). Response to comment on "high-resolution global maps of 21st-century forest cover change". *Science*, 344(6187), 981.

Huisingh, D and Z Zhang (2019). Big data mining approaches for helping to reduce and prevent global warming. *International Journal of Big Data Mining for Global Warming*, 01(01), 1901001.

Iliadis, L, V-D Anezakis, K Demertzis and S Spartalis (2018). Hybrid soft computing for atmospheric pollution-climate change data mining. In *Transactions on Computational Collective Intelligence XXX*, N Thanh Nguyen and R Kowalczyk (eds.), Lecture Notes in

Computer Science, Vol. 11120, pp. 152–177. Cham: Springer. doi: 10.1007/978-3-319-99810-7_8.

Kussul, N, M Lavreniuk, S Skakun and A Shelestov (2017). Deep learning classification of land cover and crop types using remote sensing data. *IEEE Geoscience and Remote Sensing Letters*, 14(5), 778–782.

Law, K (2012). Air pollution and climate change. Special Thematic Issue, November, Science for Environment Policy, DG Environment News Alert Service, European Commission.

Liu, X, F Pei, Y Wen, X Li, S Wang, C Wu, Y Cai, J Wu, J Chen, K Feng, J Liu, K Hubacek, SJ Davis, W Yuan, L Yu and Z Liu (2019). Global urban expansion offsets climate-driven increases in terrestrial net primary productivity. *Nature Communications*, 10(1), 5558. doi: 10.1038/s41467-019-13462.

Makantasis, K, K Karantzalos, A Doulamis and N Doulamis (2015). Deep supervised learning for hyperspectral data classification through convolutional neural networks. In *Proceedings of the 2015 IEEE International Geoscience and Remote Sensing Symposium*.

McClanahan, TR and EJ Cinner (2010). *Adapting to a Changing Environment: Confronting the Consequences of Climate Change*. New York, NY: Oxford University Press.

Mou, L, P Ghamisi and X Zhu (2018). Unsupervised spectral-spatial feature learning via deep residual conv–deconv network for hyperspectral image classification. *IEEE Transactions on Geoscience and Remote Sensing*, 56(1), 391–406.

O'Brien, M (2014). Climate change: Evidence and causes. An Overview from the Royal Society and the US National Academy of Sciences. Available at http://dels.nas.edu/resources/static-assets/exec-office-other/climate-change-full.pdf. Accessed on 11 January 2021.

Petersson, H, D Gustafsson and D Bergstrom (2016). Hyperspectral image analysis using deep learning — A review. In *Proceedings of the 2016 Sixth International Conference on Image Processing Theory, Tools and Applications (IPTA)*, pp. 1–6. Piscataway: IEEE.

Qu, T, L Wang, J Yu, J Yan, G Xu, M Li, C Cheng, K Hou and B Chen (2020). STGI: A spatio-temporal grid index model for marine big data. *Big Earth Data*, 4(4), 435–450. doi: 10.1080/20964471.2020.1844933.

Rahel, F and JD Olden (2008). Assessing the effects of climate change on aquatic invasive species. *Conservation Biology: The Journal of the Society for Conservation Biology*, 22(3), 521–533.

Seinfeld, JH and SN Pandis (1998). *Atmospheric Chemistry and Physics: From Air Pollution to Climate Change*. New York: John Wiley & Sons.

Tao, C, H Pan, Y Li and Z Zou (2015). Unsupervised spectral-spatial feature learning with stacked sparse autoencoder for hyperspectral imagery classification. *IEEE Geoscience and Remote Sensing Letters*, 8(6), 2381–2392.

Valafar, F (2002). Pattern recognition techniques in microarray data analysis. *Annals of the New York Academy of Sciences*, 980(1), 41–64. doi: 10.1111/j.1749-6632.2002.tb04888.

Wang, W, Y Li, X Luo and S Xie (2020). Ocean image data augmentation in the USV virtual training scene. *Big Earth Data*, 4(4), 451–463. doi: 10.1080/20964471.2020.1780096.

Zhou, G, X Wang, W Chen, X Li and Z Chen (2020). Realization and application of geological cloud platform. *Big Earth Data*, 4(4), 464–478. doi: 10.1080/20964471.2020.1820175.

Ziska, LH and JS Dukes (eds.) (2014). *Invasive Species and Global Climate Change*. Oxon, UK: CAB International.