

POLITECNICO DI TORINO

Master's Degree in ICT for Smart Societies



Master's Degree Thesis

**Introducing Machine Learning Solutions
For Modern Mobile Network
Management**

Supervisors

Candidate

Prof. Michela MEO

Annalisa PELATI

Prof. Paolo DINI

Academic Year 2018/2019

Abstract

Nowadays, the increasingly innovative devices and the success of social networks, together with the little tolerance of users for network bottlenecks, lead to confront with highly complex mobile network architectures. This makes difficult to monitor and manage the multitude of components ensuring satisfactory performances. To address the issue, both the cloud and edge side of mobile networks are becoming more and more sophisticated in collecting and mining the huge amounts of mobile data that users produce and consume every day. Moreover, the environmental concerns that are emerging lately, suggest the development of sustainable solutions in the management of mobile networks, enabled by the technological advancement. In this context, deep learning opens up entirely new possibilities through the systematic mining of collectable information provided by different sources.

This thesis is the result of the research conducted at the Centre Tecnològic Telecomunicacions Catalunya (CTTC) in Castelldefels. With the common objective of exploiting the Long Short Term Memory (LSTM) neural networks capability to learn the temporal context of input sequences to make better predictions, this work develops along two directions.

On the one side, it aims to define an evaluation strategy for the correct sizing of LSTM networks using as testing ground the Renewable Energy Sources (RES) forecasting; to move toward the improvement of the energy efficiency of cellular networks. The so defined LSTM architectures represent a trade-off between the accuracy of the prediction on the test set and the complexity of the network, producing good results in Wind Speed (*WS*) and Global Horizontal Irradiance (*GHI*) forecasting.

On the other side, we exploit the mobile network as an additional sensing platform, using the information exchanged over time by the different network elements (e.g., base stations, mobile terminals) for contextual Anomalies Detection (AD) purpose. It has been developed a semi-supervised algorithm, based on LSTM networks, that is taught to detect traffic anomalies learning only from non-anomalous examples individuated by different unsupervised techniques (namely, K-means and DBSCAN). The obtained results show the capability of the proposed procedure to recognize events related to know occurrences.

The challenges and the approaches addressed in this work are only few of the new solutions still to be investigated: today's unprecedented technological advance is a challenge precisely because involves major management problems and environmental impact but at the same time gives us the tools to face and solve them.

Table of Contents

List of Tables	5
List of Figures	7
1 Introduction	10
1.1 Motivation	13
1.2 State Of Art	16
2 Renewable Energy Sources Prediction for Power Supply of Base Station	20
2.1 Introduction	20
2.2 Dataset	22
2.3 Deep Learning For Time-Series Forecasting	24
2.3.1 Architecture Definition	26
2.4 Results Analysis	28
2.5 Conclusion	32
3 Urban Anomalies Identification	34
3.1 Introduction	34
3.2 Dataset	36
3.3 Anomaly Detection Framework	38

3.3.1	Data Pre-processing through Unsupervised Learning	40
3.3.2	Algorithm learning	45
3.4	Results Analysis	47
3.4.1	Unsupervised Approach Results Analysis	48
3.4.2	Semi-supervised Approach Results Analysis	50
3.5	Conclusion	54
4	Conclusion	57

List of Tables

2.1	Best LSTM configurations.	31
3.1	K parameter for the variables of interest.	42
3.2	Performances of the different AD approaches evaluated on the <i>Fiesta de San Cayetano</i> week (week 6).	50

List of Figures

1.1	Cisco VNI Global IP Traffic Forecast 2017,2022.	10
1.2	Growth in the number of BSs (©GSM Association 1999–2015).	11
2.1	Temporal variation of <i>WS</i> and <i>GHI</i>	23
2.2	Sliding windows representation.	24
2.3	Structure of a basic LSTM Memory cell.	25
2.4	Architecture of a general neural network.	26
2.5	PACF for <i>WS</i> time-series.	29
2.6	MSE related to parameters variation for <i>GHI</i> metric.	30
2.7	Comparison between the real value of 300 random samples of the test set and their predicted values for <i>WS</i> metric.	31
3.1	Scenario.	36
3.2	Temporal behavior of our variable of interest.	37
3.3	Pearson correlation plot.	38
3.4	Semi-supervised LSTM-AD Framework.	39
3.5	nRNTI <i>elbow</i> definition.	42
3.6	nRNTI number of outliers for different values of K.	42
3.7	nRNTI <i>eps</i> definition.	43

3.8	nRNTI outliers (black markers) defined through K-means unsupervised approach.	44
3.9	nRNTI outliers (black markers) defined through DBSCAN unsupervised approach.	44
3.10	Aerial view of the Rastro district, in Madrid. In red is shown the eNodeB and in green the position of the scheduled events mentioned on the festival program.	48
3.11	Period of interest individuated during the sixth week used as common test set. In green those ones related to the <i>Fiesta de San Cayetano</i> and in blue the one related to the <i>Rastro Market</i>	48
3.12	Comparison between outliers period identified by K-means and DBSCAN based semi-supervised approach.	50
3.13	p definition for DBSCAN-based semi-supervised approach.	52
3.14	Anomalies identified by DBSCAN (red point) and by the semi-supervised LSTM-based procedure (yellow circles) for different values of p	53

Chapter 1

Introduction

Mobile communication is one of the most successful innovations in modern history. During the last decade, we witnessed the huge growth in the cellular networks' market, fostered by the introduction of increasingly innovative devices and the success of social networks. The number of subscribers and the demand for cellular traffic is grown dramatically, requiring the development of data-oriented services that include, but are not limited to, multimedia communication, online gaming, and high-quality video streaming. According to *Cisco* estimations [1], the sum of all forms of IP video will continue to be in the range of 80 to 90 percent of total IP traffic, accounting for 82 percent of traffic by 2022 (Figure 1.1).

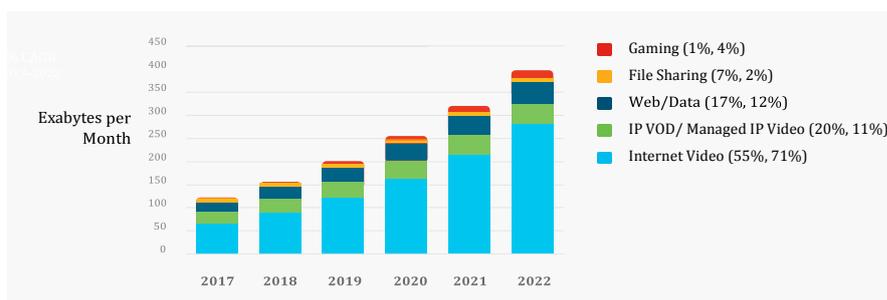


Figure 1.1: Cisco VNI Global IP Traffic Forecast 2017,2022.

The implications of this growth are difficult to estimate, but it is carrying the Internet traffic to evolve from a relatively steady stream of traffic to a more dynamic traffic pattern. Concurrently, users have little tolerance for network bottlenecks and expect high performance from applications. In this context, mobile operators have to install additional cellular base stations (BSs) to face the new demands in wireless cellular networks, trying to keep their costs at minimum. Relevant data (Figure 1.2) shows how between 2007 and 2012 the number of global BSs has doubled, exceeding more than four millions today [2]. Indeed, the latest industry forecasts indicate that the annual worldwide IP traffic consumption will reach 3.3 zettabytes (1015 MB) by 2021, with smartphone traffic exceeding PC traffic by the same year [1].

Considering that the BSs account for around 57% of the total consumed energy in cellular networks, the consequence of their growth has been a significant increase in energy consumption; producing high costs and parallel pollutant effects.

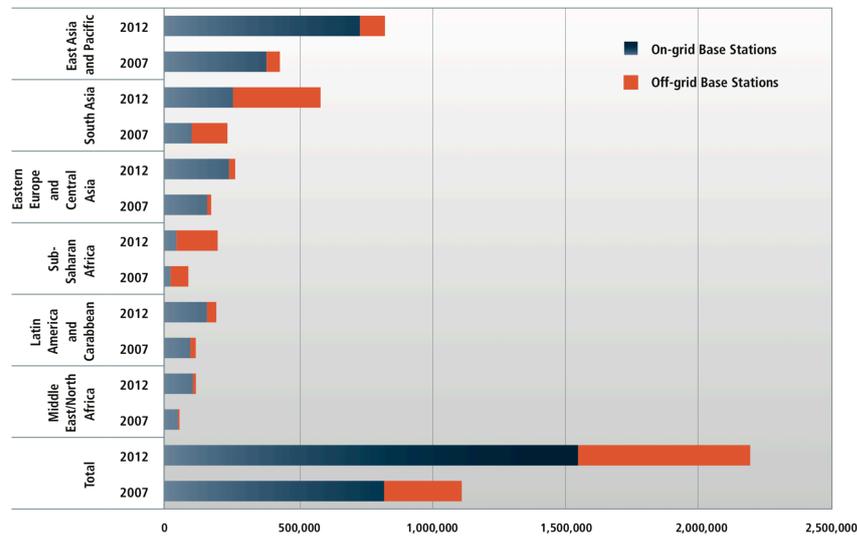


Figure 1.2: Growth in the number of BSs (©GSM Association 1999–2015).

The growing complexity and diversity of mobile network architectures have made difficult to monitor efficiently the multitude of network elements and to make sure to provide satisfactory network performance. To address the issue, both the cloud and edge side of mobile networks are becoming increasingly sophisticated to meet their users, who produce and consume huge amounts of mobile data every day. Appropriately collecting and mining these data can benefit multidisciplinary research fields and the industry, in areas such as social analysis, mobile network management, personal services provision, public transportation, and so on [3].

Therefore, introducing Machine Learning (ML) solutions into future mobile networks draws unique research interest [4, 5]. ML opens up entirely new possibilities through the systematic mining of collectable information from traffic data, and it allows to automatically discover correlations normally too complex to be extracted by human experts [6].

These assumptions are the core principles of the work carried out during the six months of research at the Centre Tecnològic Telecomunicacions Catalunya (CTTC) in Castelldefels. Among the different ML techniques and their wide range of applications in mobile networks, we decide to focus on the potentiality of deep learning approaches proceeding in two different domains/*directions*:

- The recently introduced green communication initiative, that aims to improve the energy efficiency of cellular networks not only to maintain their profitability but also to reduce the negative environmental effects of their operations.
- The usage of the mobile network as an additional sensing platform to monitor large metropolitan areas, through the detection of critical anomalies in the information exchanged over time by the different network elements (e.g., base stations, mobile terminals).

1.1 Motivations

ML and mobile networking have been researched mostly independently. Only recently crossovers between the two areas have emerged. Considering that ML is a continuously developing practice, several are the approaches with which this integration can take place [7]. Among the different techniques, we decide to exploit the deep learning benefits through the employment of the Long Short-Term Memory (LSTM) networks.

In general, deep learning achieves great performances in various data analysis tasks, on both structured and unstructured data and it performs remarkably well in feature extraction from raw data. The result is a growing interest in using these methodologies for mobile network-level and app-level data analysis. The aim is to optimize the mobile network configurations through the extraction of useful information from data collected within the network, and to perform app-level data mining thus greatly benefiting users themselves, mobile operators, and indirectly device manufacturers. At the same time, deep learning enables to understand the movement patterns of mobile users, either at group or individual level, for epidemiology study, urban planning, public service provisioning, and mobile network resource management. Also modern cybersecurity systems benefit from deep learning since it can be used to learn signatures and patterns from experience, generalizing to future intrusions, or to identify patterns that are clearly differed from regular behavior. [7]

In the multitude of possible fields of application, we decide to focus on the challenges and benefits related to the pervasiveness of the mobile networks and the ever-increasing range of services that new technologies offer to us. Today's unprecedented technological advance involves major management problems and environmental impact that cannot and must no longer be ignored. It is essential that new technologies and the enormous amount of available data are used in

the effort to deepen people’s knowledge of habits; so as to use this consciousness to develop mobile networks able to support the traffic demand with sustainable solutions.

In the wake of the already introduced growth in the number of existing BSs, in Chapter 2 we exploit the potentiality of LSTM networks in forecasting the Global Horizontal Irradiation (*GHI*) and the Wind Speed (*WS*) with respect to less complex deep learning solutions as Feedforward Neural Networks (FNNs). The attempt is to enhance the integration of renewable power generation into the traditional power network. This effort is justified by the fact that the power supply requirements for BSs, including cost-effectiveness, efficiency and reliability, can be satisfied using the technological advancements in renewable energy sources (RES) [8, 9]. An example is the renewable-based cellular BSs, that represent an ideal solution also for those areas without a mature electric network and in developed countries that suffer from continuous power cuts [10].

Despite the potentiality of this technology, the sustainability of power sources is a big challenge because service outages and power shortages are strictly prohibited in the cellular mobile sector. In order to make available fully or partially RES based solutions in those areas in which the resources are subject to seasonal variation, it is necessary to know their availability throughout the year. In this sense, we study the potential of LSTM technique in forecasting the solar irradiation and the wind speed using information concerning the city of Castelldefels, in Spain.

To provide solutions that can really be adopted in the migration of mobile networks to sustainable solutions, but more generally for assuring the full sharing of the benefits of new technologies, it is necessary to perform targeted analysis of user habits and needs. On the one side, because sustainable systems as complex as cellular networks cannot be modeled to sustain endless traffic demand. Rather, they must be integrated with solutions that allow the automatic detection of urban

anomalies, such as abnormal movements of crowds, in their early stage; so that to be able to take precautionary measures, without resulting in outages and malfunctions of the system. On the other side, because the next generation systems have the ability to be dynamically reconfigurable, driven by analytics and intelligence. In this sense, the study of mobile network users' habits, enables automatic operations that may address a multitude of issues, like capacity planning, outage detection and energy saving; in the effort to move toward the optimization of the network functioning, simplifying the maintenance and saving costs.

In Chapter 3 we exploit the extreme ubiquity of the mobile telecommunication sector to perform contextual Anomaly Detection (AD), using the information exchanged by the different network elements as pictures taken by 'panoramic cameras'; which provide a city-scale sensing system to monitor large metropolitan areas. This kind of data usually exhibits significant spatio-temporal variations resulting from users' behaviors, which can be used for network diagnosis and management, user mobility analysis and public transportation planning. We can learn from past experience by processing historically collected data, to be able to identify whether an event can be considered as anomalous. We develop an LSTM-based semi-supervised procedure to recognize complex patterns from mobile traffic and identify anomalous events automatically. The obtained results are confronted with the one reached through another ML techniques, namely K-means. To do so, we focus on data collected by an eNodeB in the Rastro district of the city of Madrid that presents *contextual anomalies*, which are classified not only by their absolute values but also based on a specific temporal context.

1.2 State Of Art

As already emphasised, the growing interest in developing a sustainable mobile network in response of today’s environmental concerns, has made the RES power forecasting an highly investigated research topic, that is rapidly evolving.

We can distinguish different types of power load forecasting models based on time horizon, algorithms and architecture types [11]. A possible way to categorize them is by their methodological foundations. A typical categorization would thereby identify physical models such as Numerical Weather Prediction (NWP) combined with solar or wind power curves, ML models such as artificial neural networks, and statistical models such the Autoregressive Integrated Moving Average (ARIMA) model. In this sense, an excellent survey of works related to solar power forecasting can be found in [12], while the area of short-term forecasting for wind speed is covered in [13] and [14]. With growing interest in possible Artificial Intelligence (AI) applications, the properties of the deep learning methodologies made it a popular method to improve the performance of load forecasting by focusing on pre-training and parameter optimization. The applied architectures range from simple FNNs [15] to more complex networks such as Recurrent Neural Networks (RNNs) [16] or time delay neural networks [17]. While FNN is the most widely applied algorithm in the literature, in [18] long short term memory (LSTM) has been used for a comparative study with the aim to forecast day-ahead solar irradiance. The results demonstrate that the LSTM networks performs significantly better than many other models (including FNNs). Similarly [19] describes a prediction method for *WS* that uses LSTM and a one-dimensional Convolutional Neural Network (CNN) fed by past information about wind speed; that is observed as time series data. The prediction accuracy and time delay, compared with the FNN’s ones are found to be improved by using LSTM and the one-dimensional CNN.

In [7] it is presented an exhaustive survey of the crossovers between the mobile and wireless networking researches and the deep learning domain, focusing on nine specific domains where deep learning has made advances. Among them we emphasise the network state prediction topic, which refers to inferring mobile network traffic or performance indicators given historical measurements or related data. A survey of the techniques recently applied to AD problems within diverse research areas and application domains is provided in [20], that presents solutions based on the mixture of traditional approaches and deep learning. In particular different techniques have been used to perform the identification of relevant features in network-level mobile data, such as mobile traffic, to obtain a characterization of network traffic behaviours. Usually this information involves essential spatio-temporal correlations that can be effectively managed by CNNs, RNNs and Autoencoders (AEs), as they are specialized in modeling spatial and temporal data. As example, in [21], it has been proposed an Autoencoder-based architecture combined with LSTMs, to model spatial and temporal correlations of mobile traffic distribution; providing better performance over SVM and the ARIMA model. Mobile traffic forecasting is performed using CNNs and LSTMs also in [22], [23] and [24], where the authors proposals gain significantly higher accuracy extracting spatio-temporal features, than traditional approaches such as ARIMA. Generally the learning procedure for AD purpose is done with supervised modality, where the algorithms are trained using a labeled dataset containing both classes (normal and anomalies) of traffic instances. In this sense, and closer to what we have done in our work, in [25] has been used a supervised approach to train a LSTM-based classifier with data captured from the unencrypted LTE Physical Downlink Control Channel (PDCCH) to identify crowded events known a-priori. In our work, instead, we use a *semi-supervised* approach to train a LSTM neural network and detect the contextual traffic anomalies associated to different urban

events. In such a way, we overcome the so called *unbalanced class problem* [26], where one class is poorly represented with respect to the other. As a result, the AD problem is not addressed as a supervised classification task, but rather, our algorithm is taught to detect traffic anomalies learning only from non-anomalous examples.

Chapter 2

Renewable Energy Sources Prediction for Power Supply of Base Station

2.1 Introduction

Over the last decade, we observed an unprecedented growth in renewable energy production. On this trend solar and wind took on new prominence, becoming the major renewable energy sources (RES) because of their great availability and accessibility. According to IEA projects, by 2050 around 15-18% of global electricity will be generated from wind, and the solar photovoltaic will contribute up to 16 % [27]. The usage of renewable energies in the mobile communications sector is not a novelty, and it will proliferate because of its capacity to be suitable to developing countries, as well as to rural and remote areas [10]. This is true mainly for solar, because of its simplicity in deployment compared to other sources, especially for low loads such as BSs in urban areas.

Nowadays, the grown number in mobile subscribers and the increased number of BSs has significantly affected the energy consumption, since these stations account for approximately 57% of the consumed energy in cellular networks. The energy efficiency turns out to be a fundamental requirement for cellular network operators not only to maintain their profitability but also to reduce the negative environmental effects of their operations. According to [28], the amount of carbon dioxide (CO₂) emitted by the mobile sector will reach 179 MtCO by 2020 and account for 51% of the total carbon footprint of the Information and Communication Technologies (ICT) sector. Therefore, the usage of renewable energy for powering cellular BSs, reducing the operational expenditures (OPEX) of cellular networks, and diminishing the greenhouse gasses (GHG) emissions have become of greater interest, representing an ideal long-term solution for the mobile cellular network industry.

Despite its great attractions, the increasing renewable power generation creates new challenges to be addressed; not only for the integration measures to be adopted, but also for the intermittent nature of the resources which leads to large variability and uncertainty in the power output.

Therefore, renewable energy systems must be integrated with non-renewable energy sources and/or modality for energy storage to ensure energy supplies and to improve system reliability, preventing mobile service outages [9, 10] that have very high societal costs.

The cellular BS can indeed be powered directly from the RES, through an autonomous or hybrid design with other means of renewable or non-renewable energy. For optimal performance of hybrid solar PV/wind systems, the implicit variability of these resources should be studied and modeled through the use of forecasts; to reduce the energy balancing, to schedule the power generation and to dispatch optimally realised decisions [9]. The primary purpose of the

forecasting procedure is to determine as accurately as possible the power output of the generation plants in the near term (15-minute, 30-minute or hour-ahead) and day-ahead time periods. Deep learning NNs, with their capacity to automatically learn arbitrary complex mappings from inputs to outputs and support multiple inputs and outputs, results to be an efficient way to approach the problem.

We apply deep learning methodologies, and more in detail Long-Short Term Memory neural networks (LSTM) to perform RES forecasting, with quarter granularity. The information that we used are provided by the *HelioClim3* web-service and related to the city of Castelldefels, in Spain. We proceed by varying the LSTM network's parameters to decrease as much as possible the RES prediction error, confronting the results with the ones obtained with simpler Feedforward neural networks (FNNs).

The chapter is organized as follows: in Section 2.2 we investigate the nature of the sources' peculiarities used for our scope, i.e. the Wind Speed (*WS*) and the Global Horizontal Irradiance (*GHI*). In the following Section, we present and justify the approach used for their forecasting purpose, presenting in Section 2.4 the obtained results.

2.2 Dataset

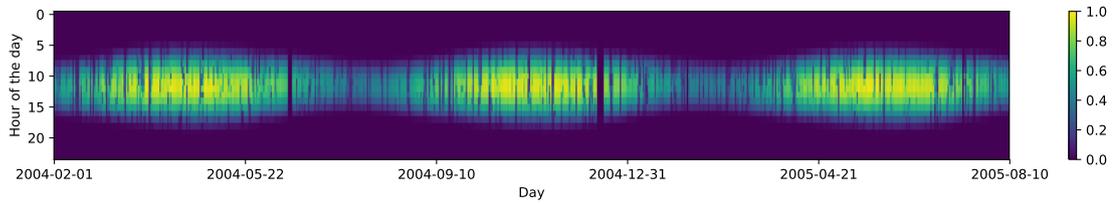
The dataset that we used for RES forecasting is provided online by *SoDa HelioClim3*, a web service for professionals of solar energy, that contains freely accessible worldwide climatic reports, from which we extrapolate information about the city of Castelldefels, Spain from February 2004 to October 2006.

SoDa HelioClim3 provides many services of solar data resources that can be valorized in companies' activities for the estimation of profitability of photovoltaic prospects projects, to monitor the solar powerplants performances and to allows the management of electricity sales on the market from solar forecast data [29].

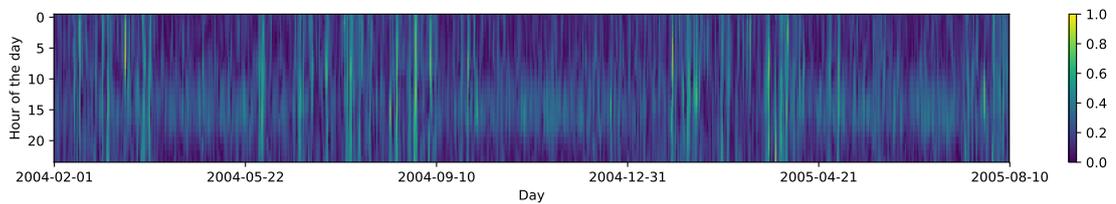
The values are estimated from satellite imagery, on a geographical area composed by Europe, Africa, Atlantic Ocean and Middle East, and data are available with a time step ranging from 15 min to 1 month.

Among the numerous information provided by *SoDa HelioClim3* we extrapolate the time-series related to *WS* and *GHI*, that represents the total solar radiation incident on a horizontal surface.

Figure 2.1 shows how the time-series concerning the *GHI* present inter-seasonal variations due to the changing day length and makes evident the associated changes in solar altitude with time of year. The *WS* time-series, on the other hand, has a variable trend and is not subject to any hourly/seasonal dependency.



(a) *GHI*.



(b) *WS*.

Figure 2.1: Temporal variation of *WS* and *GHI*.

The lack of correlation of the two variables is confirmed through the evaluation of the Pearson Correlation between them, which results lower than the 4%.

For the forecasting purpose, the dataset has been normalized with respect to the peak values that occurred in the examined period and, before deep learning can be used, time-series must be appropriately re-framed in a form that is understandable

to a mathematical model. For this purpose a sliding window of length W is used to cut out a frame of data points that are employed to forecast the N time-steps after the end of the window (Figure 2.2).

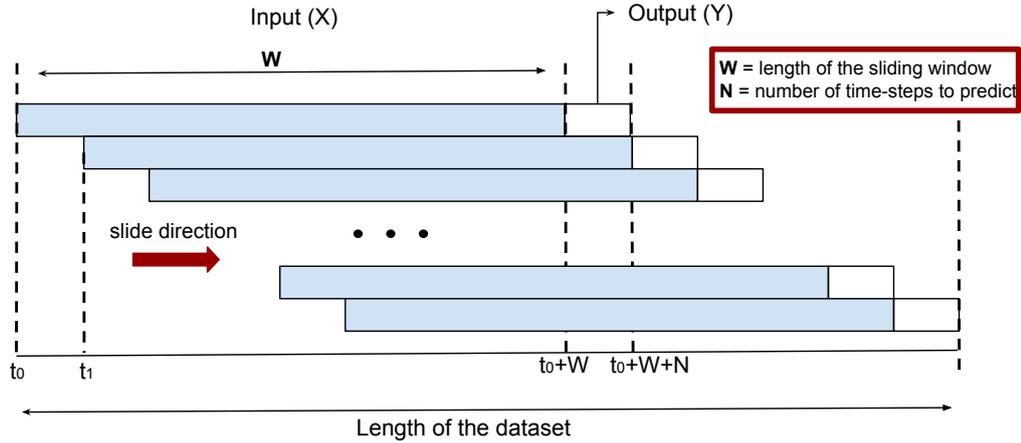


Figure 2.2: Sliding windows representation.

2.3 Deep Learning For Time-Series Forecasting

The time-series forecasting problem can be framed as a supervised learning problem. Generally, with supervised learning the models are fit on training data containing both inputs and outputs to perform predictions on test sets, where only the inputs are provided. Then the model's output is compared to the target variables to estimate the capability of the model. A wide range of supervised learning algorithms is available, each with its strengths and weaknesses.

Among the different approaches, the capacity of deep learning methodologies to automatically learn the temporal dependencies and handle the temporal structures, like trends and seasonality, make them particularly suited to solve problems like ours. Time-series data can be highly erratic and complex and deep learning methods make no assumption about their underlying pattern, are robust to noise (common

in time-series data) and provide direct support for multi-step and even multivariate forecasting.

In particular, we decide to perform RES forecasting through recurrent neural networks (RNNs), which are the state-of-the-art learning techniques to cope with sequential input data. RNNs are deep learning networks with a twist: they are composed by cells with two incoming connections, the input and the previous state, and similarly they have two outgoing connections, the output and the current state. This state helps in combining information from the past and the current input, and are thus preferred for sequence processing.

The LSTM network that we used for forecasting purpose, is a particular kind of RNN, composed by cells with ‘gates’ that represents the state input (Figure 2.3). The intuition behind these gates is to forget or retain information from the past, which allows them to remember more than just the immediate past. This kind of NN adds native support for input data composed of sequences of observations. The model learns a mapping from inputs to outputs and identifies what context from the input sequence is useful for the mapping; dynamically changing this context as needed. This capability suggests that the promise of LSTM networks is to learn the temporal context of input sequences in order to make better predictions.

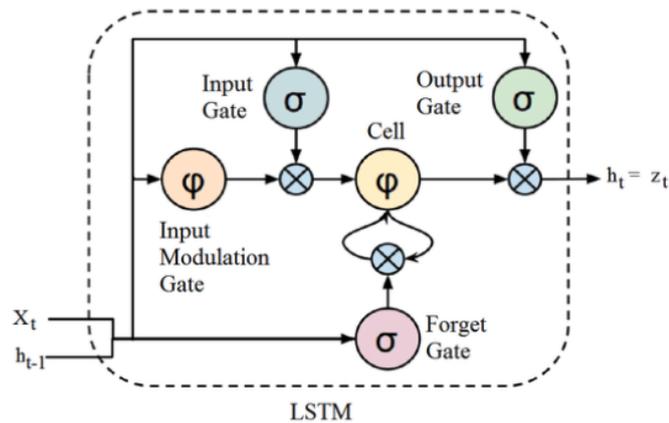


Figure 2.3: Structure of a basic LSTM Memory cell.

2.3.1 Architecture Definition

Many are the attributes that must be decided upon when designing a neural network model. Each model consists of architecture and few parameters, and for a given architecture the values of the parameters determine how accurately the model performs the task. Some of the more important architecture parameters, in terms of training and network capacity, are the number of hidden layers (nHL) and the number of cells (nC) composing each layer, to keep the networks as simple as possible without affecting the capacity of performing on input data (Figure 2.4).

Even if the usage of many layers of hidden neurons enables greater processing power and system flexibility, the cost is the additional complexity in the training algorithm. Too many hidden neurons lead to over-specified systems, incapable of generalization; as a system of equations with more equations than free variables. Having too few hidden neurons, conversely, reduces the robustness of the system, preventing a properly fitting of the input data.

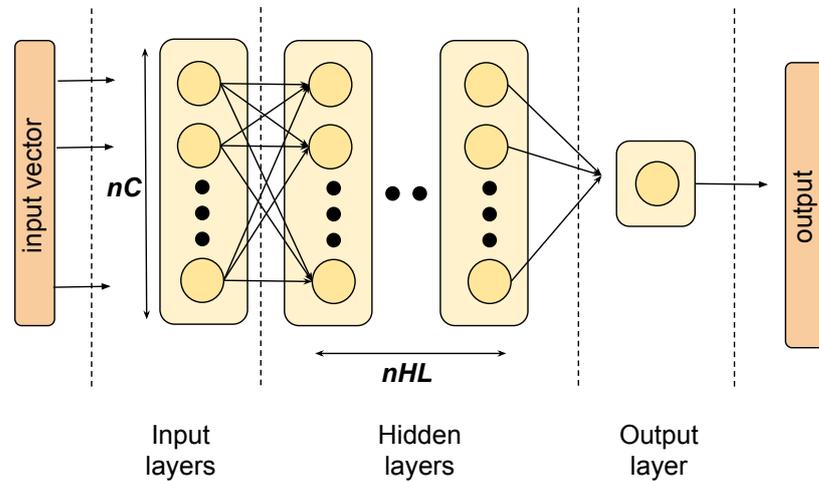


Figure 2.4: Architecture of a general neural network.

Another important characteristic that could affect the performance in a meaningful way is the fixed-length frame W used to reshape and window the input dataset.

In other words, W defines the number of time-lags that the architecture processes to make the prediction on the following time instant (Figure 2.2). This affects significantly the network training time and the effectiveness of the procedure.

Even it might sound ‘old’ in the era of self-learning machines, the best procedure to define these values is with manual attempts. For this reason, we proceed to test the MSE of the prediction for different values of nHL , nC and W both for GHI and WS variables.

In statistics, the Mean Squared Error (MSE) represents a measure of the quality of a predictor or an estimator. In a vector of n predictions generated from a samples of n data points, it is defined as the average squared difference between the predicted values (\hat{Y}_i) and the actual values (Y_i):

$$MSE = \frac{1}{n} \sum_{i=1}^n (Y_i - \hat{Y}_i)^2. \quad (2.1)$$

It is always non-negative, and values closer to zero are referred to better predictions.

The exigency of testing a high number of different combinations of parameters has prompted us to define a procedure that consists of the evaluation of the MSE variation on the prediction, proceeding in two steps. At first, we fix the nHL to 1 and we vary nC and W to establish a suitable value for them; that’s because their increase affects in a limited way, even if not null, the time needed to train the network.

Once W and nC have been appropriately defined, we evaluated the MSE varying the nHL value to find the final optimal network configuration, that represents a trade-off between the accuracy of the prediction on the test set and the complexity of the network.

2.4 Results Analysis

The LSTM performances have been evaluated in terms of MSE of the prediction using *Google Colaboratory*, which provides free hardware acceleration with *Tensor Processing Unit (TPU)*. The dataset is split into training and test set with a ratio of 80% - 20%. The forecasting algorithm has been implemented in *Python*, using *keras* library and *Tensorflow* as backend.

While the procedure we used for the optimal network sizing continue with varying nC , nHL and W , some elements remain unchanged regardless of them:

- All the LSTM networks consist of at least one LSTM layer and a final Fully Connected (FC) layer composed by a single neuron to perform the prediction;
- Both the LSTM layers and the FC layer use the *ReLU* activation function to process the output;
- The algorithm is trained using the *Mean Square Error* loss function and optimized using the *Adaptive Learning Rate SGD Algorithms (Adam)* optimizer, chosen because it is fast in terms of convergence and highly robust to model structures [30].

Although the manual method is certainly the optimal solution in the definition of the architecture parameters, the nature of the time-series with which we confronted cannot be forgotten in this delicate phase. For this reason, we defined the interval of values in which W has been varied studying the *Partial Autocorrelation Function (PACF)* of the series. In time-series analysis it gives the partial correlation of a stationary time-series with its own lagged values, regressed the values of the time-series at all shorter lags.

Looking at the PACF of the WS variable, in Figure 2.5, and trying to perform predictions using as few previous information as possible, we fix the range of possible values for W between 3 and 6. The same interval has been fixed also for the GHI metric looking at the correspondent PACF.

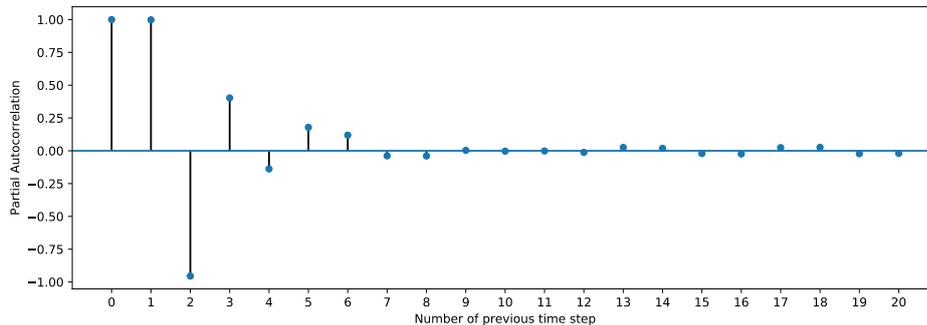


Figure 2.5: PACF for *WS* time-series.

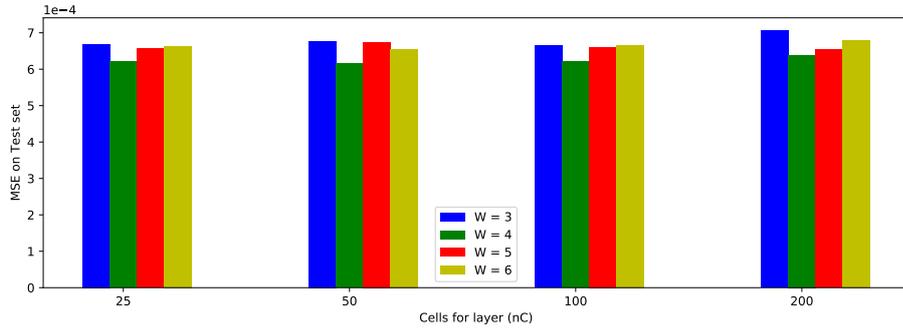
On other hands, we arbitrarily decide to vary nC to assume a value contained in the carrier $c1=[25, 50, 100, 200]$ and nHL in the carrier $c2=[1, 2, 3, 4]$.

The results obtained through the procedure based on the manual attempt considering the *GHI* variable are shown in Figure 2.6. The MSE values calculated on the test set reveal that the Nc variation do not particularly affect the performances of the network forecasts and also varying W , the MSE of the prediction remains rather stable. Therefore, nC is fixed equal to 25 and W equal to 4.

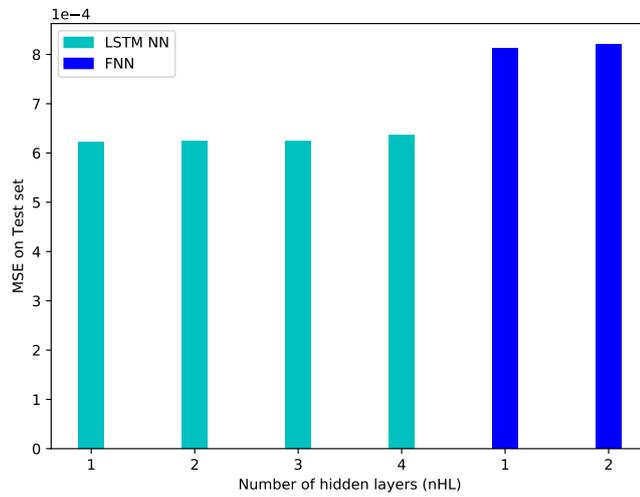
Making the same considerations, once fixed nC and W , the number of nHL is set equal to 1 in the final optimal LSTM architecture: the choice to further complicate the composition of the network is not justified by observing the low gain in performances shown in Figure (2.6.b).

To compare the performance of the optimal LSTM network with another deep learning approach, we test the capability of a Feedforward network (FNN) applied to the same problem. To provide a stable basis of comparison, nC and W are held equal to those defined for the LSTM network, while the number of FC hidden layers is varied between 1 and 2. Despite Figure 2.6.b shows that the error on prediction is slightly higher using this approach, it remains very low.

The nature of the *WS* time-series is extremely different from the *GHI*'s one; and this is underlined both by the Heatmap in Figure 2.1 and by the low correlation



(a) MSE varying nC and W with fixed $nHL = 1$.



(b) MSE varying nHL with fixed $nC = 25$ and $W=4$.

Figure 2.6: MSE related to parameters variation for GHI metric.

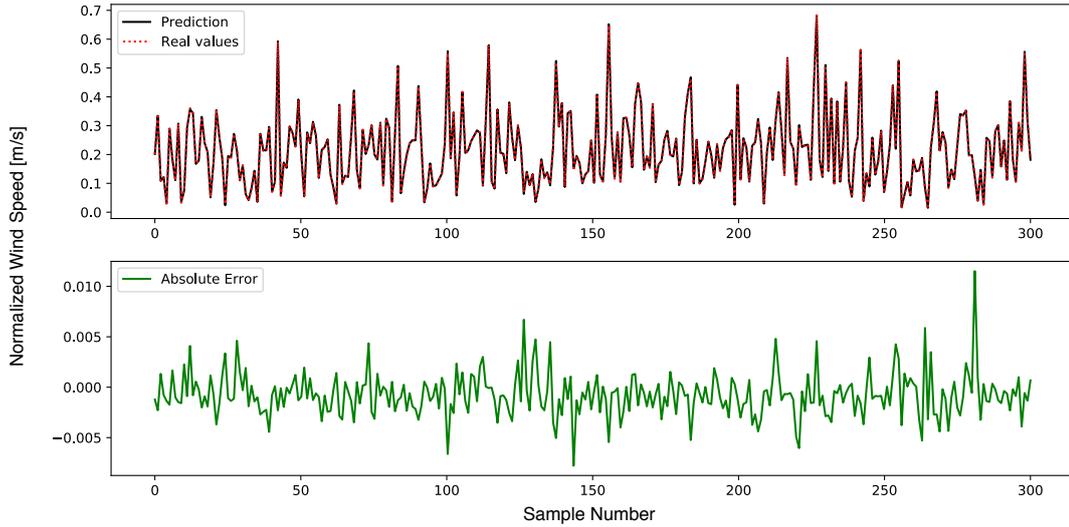
between the two metrics. Despite of this consideration, the LSTM configuration that results to be the optimal one for the WS forecasting, is the same as for the GHI forecasting.

In Table 2.1 we summarize the performance of the best LSTM network configuration for both the investigated metrics.

As prove of the high performances of the optimal LSTM network, we show in Figure 2.7 the comparison between the predicted values of WS and the respective

Table 2.1: Best LSTM configurations.

	W	nC	nHL	MSE
GHI	4	25	1	$6.217 \cdot 10^{-4}$
WS	4	25	1	$3.053 \cdot 10^{-6}$

**Figure 2.7:** Comparison between the real value of 300 random samples of the test set and their predicted values for WS metric.

real values on 300 random samples of the test set, emphasizing the absolute value of the error for each of them.

These results demonstrate that the peculiarities of the LSTM networks make them a suitable approach to forecast both metrics that varies following a more visible trend, like GHI , and metrics with very floating behavior like WS .

Although it is about rather short forward time-steps forecasting (15 minutes), we know that by their nature this kind of data are quite noisy. The quality of the obtained predictions, especially for WS , suggests that the data provided by *SoDa HelioClim3* web-site are of very high quality. It could be related to the fact that *SoDa HelioClim3* offers to companies a paid service that provides more

current data than those used for our work. This certainly cause these data to be pre-processed and cleaned so that they can be used efficiently and directly by customers.

2.5 Conclusion

To face the today's environmental concerns it is necessary to move toward the development of mobile networks able to support the traffic demand with sustainable solutions. The feasibility of such solutions are implicitly related to the ability in forecasting renewable energy sources (RES) availability, such as solar and wind, in the short and medium term. We present an evaluation strategy for the correct sizing of LSTM networks to forecast with good accuracy the RES intensity at one time-step ahead. At the end, the designed LSTM's architecture represents a trade-off between the accuracy of the prediction on the test set and the complexity of the network. For this purpose, we employ data available online through the *SoDa HelioClim3* web service, which provides climatic reports from which we extrapolate those related to Castelldefels, in Spain, with a time granularity of 15 minutes. The constructed LSTM networks perform very well in Wind Speed (*WS*) and Global Horizontal Irradiance (*GHI*) forecasting. In particular we demonstrate the higher performances of LSTM networks in learning temporal dependencies of data with respect to FNNs.

Chapter 3

Urban Anomalies Identification

3.1 Introduction

Nowadays the 55% of the world's population reside in urban areas. The future projections reveal that this estimation is expected to grow, reaching the 68% by 2050. According to United Nations data [31], the urbanization process, that is based on a gradual shift in population's residence from rural to urban areas, combined with the growth of the world's population, could increase of other 2.5 billion the urban areas population by 2050, with close to 90% of this growth in Asia and Africa.

To ensure that the benefits of urbanization are fully shared and inclusive, a sustainable development of metropolitan areas is crucial. This is straightly related to the efficient management of cities, including housing, transportation, energy systems, education and health care. In this sense, the detection of urban anomalies such as abnormal movements of crowds, is of great importance for government

and public administration [32], which could benefit even more if anomalies can be automatically alerted in their early stage.

However, urban anomalies often exhibit complicated forms. Monitoring the heterogeneous sources that may provide useful information requires complex and advanced detecting tools, which have elevated deployment and maintenance costs. To deflect this problem, we advocate the use of mobile networks as additional sensing platform used to exploit the extreme pervasiveness of the mobile telecommunication sector within the urban population and its ubiquitous coverage [33] to monitor large metropolitan areas. In this sense, the detection of critical anomalies can be achieved through the collection of information that the different network elements (e.g., base stations, mobile terminals) are exchanging over time. Moreover, learning from past experience by processing historically collected data allows to identify whether an event can be considered as anomalous.

For this purpose, we use unsupervised algorithms, namely DBSCAN and K-means, to label past data as normal or anomalous samples. The dataset composed only by the normal ones is then used to train a stacked LSTM architecture to predict the traffic at the next time-instant. Actually the LSTM networks is able to recognize complex patterns from mobile traffic information considered as time series and the Anomaly Detection (AD) is performed comparing the prediction error against the ground-truth. In this work we focus specifically on *contextual anomalies*, classified as such not only by their absolute values but also based on a specific temporal context. For example, a period of high traffic would be correctly classified as non-anomalous at regular peak times, but it would be an anomaly at low traffic hours (e.g. during the night).

The chapter is organized as follows: in Section 3.2 we investigate the nature of the information used for the AD purpose, of which we deepen the concept in Section 3.3; where we present our semi-supervised AD procedure in all its stages.

In Section 3.4 we show the obtained results, with particular attention in comparing them with the alternative ML methodology namely K-means.

3.2 Dataset

The dataset used for our work is related to LTE scheduling information collected in the Rastro district, in the centre of the city of Madrid, from the end of June, 2016 to the beginning of August, 2016. The area is surrounded by commercial activities like restaurants and shops, with a neighboring crowd related to pedestrians and vehicles. Data has been collected from the LTE Physical Downlink Control Channel (PDCCH) using an LTE sniffer, as in [34], able to decode the Downlink Control Information (DCI) message sent from the eNodeB to the connected UEs (Figure 3.1). DCI messages contain information about resources assigned in the Uplink (UL) and Downlink (DL) communications with the network, identifying each UE in RRC-mode by a C-Radio Network Temporary Identifier (RNTI).

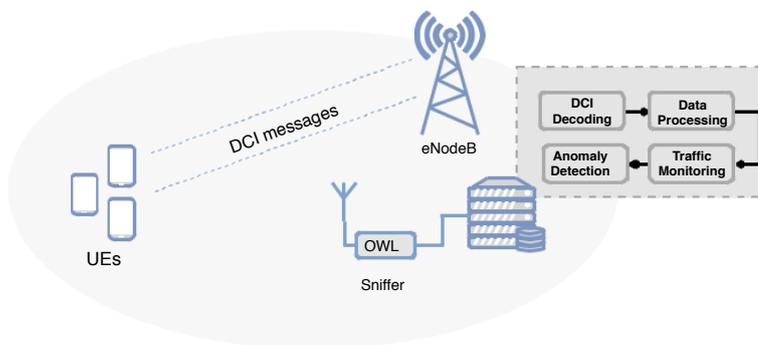


Figure 3.1: Scenario.

At each Transmission Time Interval (TTI), which is 1ms in LTE, we can access:

- The number of connected users (n_{RNTI})
- The number of allocated resource blocks, both in Uplink (RBU_p) and Dowlink

(RBDDown), where the Physical Resource Block (PRB) represents the smallest resource unit in time and frequency that can be allocated to any user.

To be able to identify traffic outliers with that scheduling information directly obtainable from the DCI message we decided to use these details as metrics for our work.

Understanding the nature of the collected data and identifying patterns is fundamental in any data-driven approach like our: the following analysis aims at describing the temporal variation of the variable of interests of the collected traces. The traffic is normalized with respect to the peak traffic that occurred in the examined period and aggregated to reach a one-minute granularity. A strong relation between the number of connected users, the hour of the day and the day itself is shown in Figure 3.2.a, that represents the average traffic profiles over 24 hours of the considered eNodeB in each day of the week.

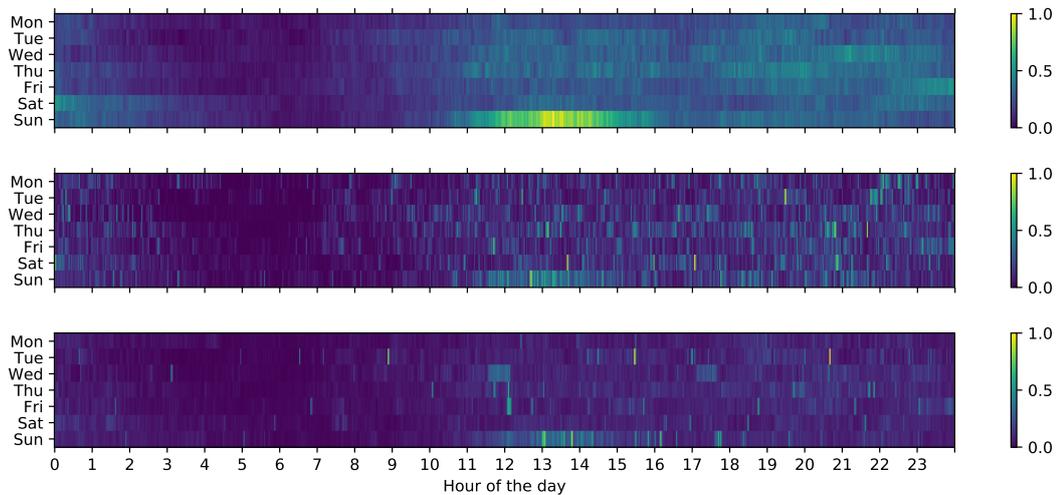


Figure 3.2: Temporal behavior of our variable of interest.

The underlined daily pattern perfectly represents people’s normal routine: high nRNTI is shown during the day (when the population is active), whereas less nRNTI is detected during nights (when people usually sleep). It is possible to

identify a different behavior on weekend days w.r.t. the weekdays. This is directly related to individuals' tendency to move their routine forward of few hours during the weekend. Moreover, a peak in the number of connected users is visible during the Sunday, which coincides with the Rastro market: a periodic flea market that takes place in this area weekly, gathering a larger number of people.

This is not true for the RBUp and RBDown features: even if a zero traffic is verified during the night (when no users are connected), it is not visible any relevant correspondence with the nRNTI. This lack of correlation is further underlined looking at the Pearson Correlation in Figure 3.3. This means that an increment in nRNTI in a cell, does not correspond to an increase in the PRBs assignment.

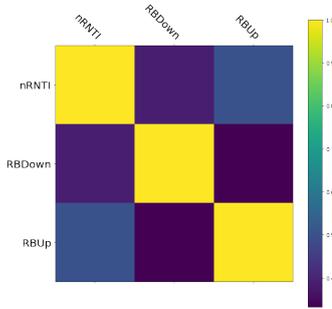


Figure 3.3: Pearson correlation plot.

On these considerations, our work analyzes the ability of different AD approaches in identifying the periods of anomaly on very different time series traffic traces, represented by our variables of interest (nRNTI, RBUp, RBDown).

3.3 Anomaly Detection Framework

The anomaly detection technique is used in datasets to automatically identify those patterns or events representing a perturbation of the expected behavior, called outliers. According to [35], there are three suitable approaches to find this kind of deviations: the unsupervised, the supervised and the semi-supervised approach.

The main difference between them is related to the prior knowledge of the data: the unsupervised methodology is better suited to those situations where no background known of the problem is possible, while the supervised one requires pre-labeled data tagged as normal or abnormal. Halfway between the two, sets the semi-supervised approach we used for our work.

We perform AD through a stacked LSTM network trained with a dataset created by windowing those points labelled as normal by two different unsupervised learning techniques: K-means and DBSCAN. The idea is to compare the capability of the pure K-means algorithm in finding the contextual anomalies with respect to the ones individuated by combining it with the LSTM potentiality. The motivation hidden behind the choice of using this type of RNN is related to the temporal nature of the available information and its promise, already introduced in 2.3, of learning the temporal context of input sequences in order to make better predictions.

To deeply analyze the potentiality of our semi-supervised approach, we test the combination between unsupervised and deep learning using also DBSCAN, because of its intrinsic difference in the clustering process with respect to K-means.

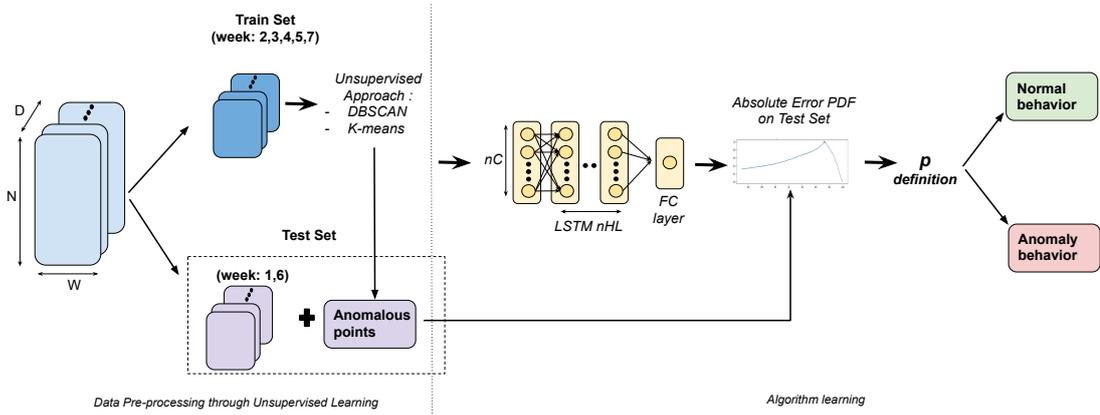


Figure 3.4: Semi-supervised LSTM-AD Framework.

A representation of the so constructed AD problem is shown in Figure 3.4. It takes as input the data collected from the LTE PDCCH and it consist of two stages: Data *Pre-processing through Unsupervised Learning* and *Algorithm Learning*. The details of each part are discussed in the next sections.

3.3.1 Data Pre-processing through Unsupervised Learning

Unsupervised learning is a branch of ML used to discover the underlying structure of the data. It is based on algorithms able to identify the presence or absence of commonalities in each piece of data without known reference.

The clustering algorithm is the well known unsupervised ML approach used for our work, able to identify classes amongst a group of objects through a measure of similarity. While various similarity measurement exists, we choose the Euclidean distance as reference: a small distance between two objects implies a strong similarity whereas a large distance implies a low similarity.

Two different typology of clustering methodology have been used, namely K-means and DBSCAN. The main difference between the two approaches is related to the clustering principles: the K-means clustering algorithm is a partition-based algorithm, while the DBSCAN algorithm is a density-based algorithm. The advantage of density-based over partition-based algorithms is related to their possibility to find clusters of arbitrary shapes and not only spherical shaped clusters as partition-based algorithms.

Both will be briefly described in the sections below.

K-means

The K-means algorithm is one of the simplest partition-based clustering algorithm. It partitions objects of a dataset into a fixed number of K disjoint subsets. For each cluster, the algorithm maximizes the homogeneity by reducing the square

error between the elements and the centre (or mean) defined as

$$E = \sum_{i=1}^K \sum_{j=1}^n |dist(x_j, c_i)|^2, \quad (3.1)$$

where x represents each object of the dataset and c the center (or mean) of each cluster.

This is performed starting from randomly chosen center of clusters and assigning all the objects of the dataset to the nearest one. K-Means iteratively computes the error of the assignment, and calculate new centers with the aim to minimize it. The procedure continues until the membership within the clusters stabilizes. At the end, for the AD purpose, all the points belonging to the less numerous cluster in the final partitioning are defined as outliers.

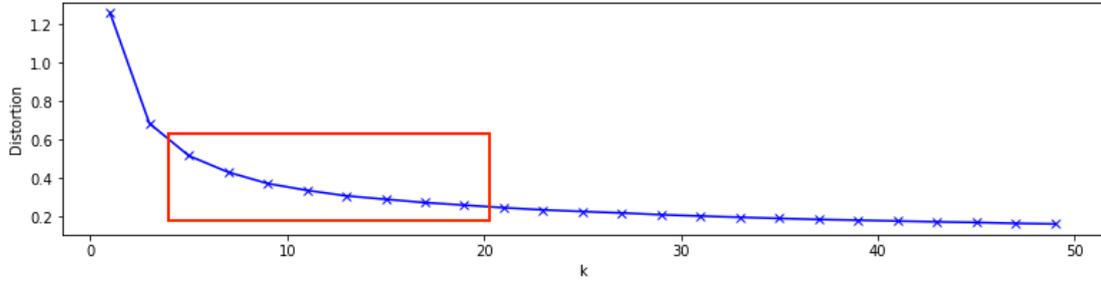
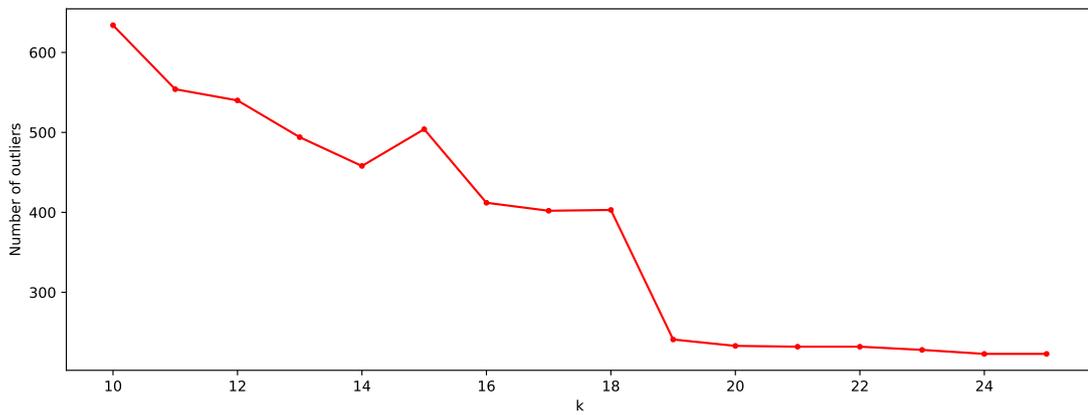
A different K parameter has been identified for each variable of interest through the Elbow method [36]. This visual method starts from low K, iteratively increasing its value to evaluate the different costs of the training. If K increases, the average distortion will decrease and each cluster will have fewer constituent instances closer to their respective centroids. However, the improvements in average distortion will decline as K increases. The value of K at which the improvement in distortion declines the most is called the *elbow*.

As example, in Figure 3.5 is shown the distortions produced by different K values for the nRNTI parameter. With our variables this approach is not able to define a unique value for K, but rather it identifies a range of possible values on which an in-depth analysis has been made.

We calculated the number of outliers identified by each value of the range, fixing K equal to the one able to stabilize it (Figure 3.6). The same procedure has been performed for all the variable of interest: Table 3.1 shows the defined K for each of them.

Table 3.1: K parameter for the variables of interest.

K	nRNTI	RBU _p	RBD _{Down}
	19	7	10

**Figure 3.5:** nRNTI *elbow* definition.**Figure 3.6:** nRNTI number of outliers for different values of K.

DBSCAN

DBSCAN (Density Based Spatial Clustering of Applications with Noise), and generally all density-based algorithm, considers clusters as dense areas of objects that are separated by less dense areas. The DBSCAN algorithm is based on the concepts of density-reachability and density-connectivity, represented with two input parameters: epsilon (*eps*) and the minimum number of points (*minPts*):

- Eps represents the minimum distance between two objects to be considered similar.
- $MinPts$ is the minimum number of points that a cluster must contain to be defined as such.

In such a way a cluster is defined as the set of objects in a data set that are density-connected to a particular core object. Their number is not decided a priori, but it changes to vary parameters. Any object that is not part of a cluster is categorized as noise and, in our case, as outlier.

According to [37], the value of $MinPts$ has been fixed equal to 4. The eps value, instead, has been defined looking at the maximum slope of the ordered vector composed by the Euclidean distances of each point to the nearest $MinPts^{\text{th}}$ point. Again, in Figure 3.7 is shown as example the nRNTI eps definition procedure.

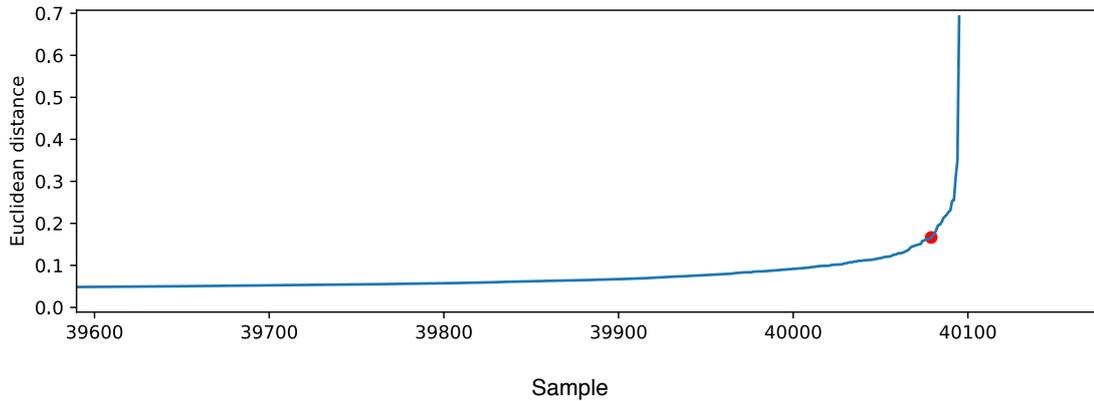


Figure 3.7: nRNTI eps definition.

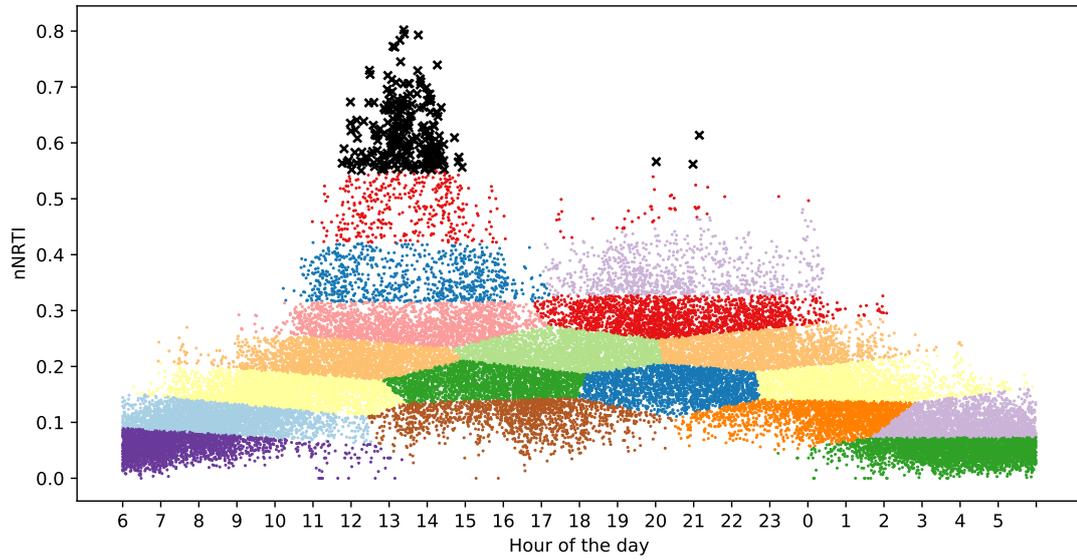


Figure 3.8: nRNTI outliers (black markers) defined through K-means unsupervised approach.

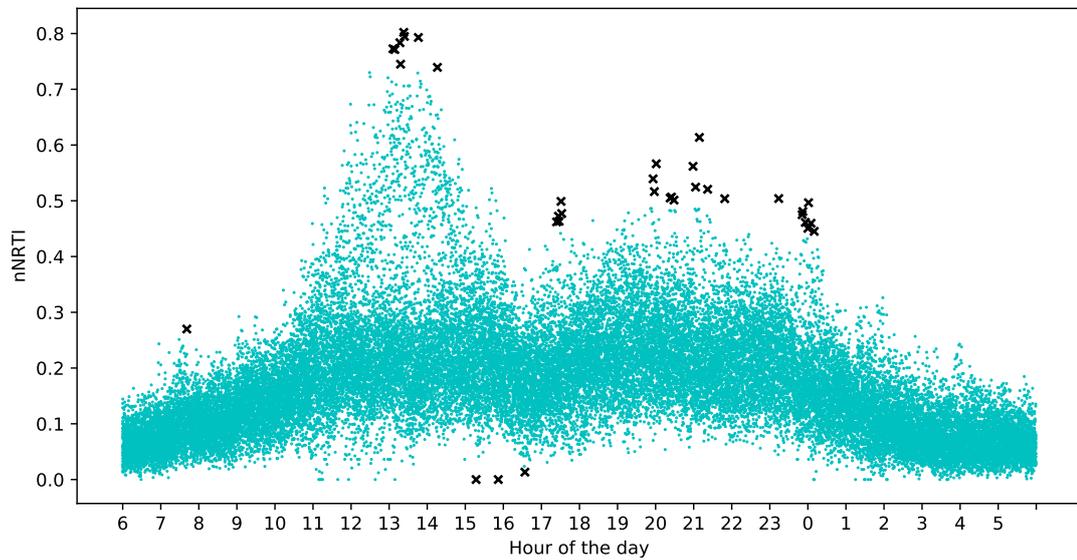


Figure 3.9: nRNTI outliers (black markers) defined through DBSCAN unsupervised approach.

3.3.2 Algorithm learning

The architecture adopted for the AD semi-supervised procedure is based on LSTM neural networks. Our choice is related to their ability of learning long-term dependencies, thanks to the structure of the basic LSTM cells (or units) that includes special gates to regulate the learning process. LSTM network is able to keep the contextual information of inputs by integrating a loop to flow the information from one step to the following one, presenting itself as an optimal solution for the analysis of time series such as our (2.3). In our design we consider a stacked architecture combining 2 LSTM hidden layers (nHL), each with 50 LSTM cells (nC) and a final Fully Connected (FC) layer composed by a single neuron to perform the prediction. The length of the observation window W is equal to 5 and it is equivalent to the number of lags of the stacked LSTM architecture.

The parameters choice (nHL , nC , and W) has been done through the same procedure used in Section 2.4 and it represents a trade-off between the accuracy of the prediction on the test set and the complexity of the network. While the MSE on the test set remains quite stable with respect to the parameters' variation, the time needed in the training phase increase with the growing complexity of the network. This justify our choice not to further complicate the network design for a low gain in performance.

Both the LSTM layers and the FC layer use the *ReLU* activation function to process the output. The algorithm is trained using the *Mean Square Error* loss function and optimized using the *Adam* optimizer.

The prediction obtained with this approach is nothing but a value that can be more or less different from the real one, for each point of the Test set. The idea behind this procedure is to train the algorithm using the samples tagged as normal by the unsupervised techniques, producing greater prediction error on the abnormal ones. Based on that, we compare the predicted values with the expected ones, to

define a measure of the prediction's absolute error for each element of the test set. The Probability Density Function (PDF) of the absolute error is used to identify the outliers, by setting a threshold that represents the boundary value behind which the prediction error is considered too high to not be referred to an anomalous point.

We decide to fix the threshold, called p , measuring the test's accuracy through the F-score metric, defined as the harmonic mean of precision (P) and recall (R) of the prediction:

$$F = 2 \frac{RP}{R + P}, \quad (3.2)$$

where:

- P represents the ratio between the number of anomalous samples that are correctly classified (true positive, T_p), and the sum between T_p and those normal samples that are incorrectly classified as anomalous (false positive, F_p),

$$P = \frac{T_p}{T_p + F_p}. \quad (3.3)$$

- R (known as hit-rate or sensitivity) represents the ratio between the T_p , and the sum between them and the number of anomalous samples incorrectly classified as normal (false negatives, F_n),

$$R = \frac{T_p}{T_p + F_n}. \quad (3.4)$$

Intuitively P represents the ability of the system not to label as anomalous a sample that is normal, and R intuitively represents the ability of the system to find all the anomalous points.

We compare the vector of the anomalous points detected by the unsupervised techniques with the one obtained by the predictive algorithm when changing the parameter p . At the end, p is fixed equal to the value able to maximize F .

A visual representation of the procedure is given in Section 3.4.2.

3.4 Results Analysis

Both the supervised and the semi-supervised methodologies have been tested considering the occurrences in the eNodeB coverage area during our period of interest. We know that the *Fiesta de San Cayetano* took place during the sixth week of the observation period, and that each Sunday the *Rastro Market* take place from 9 to 15. We use such knowledge as common basis to evaluate the performances of the algorithms, by verifying the capacity of the different AD procedures in identifying the events related to the occurrences during the sixth week.

Figure 3.10 shows in red the position of the eNodeB and in green the position of the scheduled events mentioned on the festival program, available online [38]; and Figure 3.11 shows the relevant time intervals: in green those ones related to the *Fiesta de San Cayetano* and in blue the one related to the *Rastro Market*. The metrics chosen for the evaluation of the performances are the well known F , P and R (Section 3.3.2).

nRNTI metric (visible in Figure 3.2). The anomalous points identified by the two methodologies turn out to be extremely different.

Figure 3.8, in which the outliers are represented by black markers, shows how the K-means approach applied on the nRNTI metric seems to identify a boundary value above which all the points are considered belonging to the anomalous cluster, identifying the majority of the outliers in the interval 11-15. Instead, DBSCAN evaluates the density of the points and their position, identifying as abnormal only the points isolated concerning to the others (Figure 3.9). It finds only few anomalies between the 11 and the 15, while it identifies other outliers between the 17 and midnight that K-means ignores.

We check the potentiality of the pure K-means approach in identifying the contextual anomalies using the test set we introduced before. Table 3.2 shows how the merely unsupervised approach is not enough to correctly find our periods of interest using any of the evaluated metrics: the F values results to be very low because negatively influenced from the R values. This is related to the inability of the pure K-means approach in finding most of the abnormal points. Figure 3.12 supports these results showing how the contextual anomalies identified by the pure K-means approach using the nRNTI variable are few and of short duration.

Since DBSCAN works by evaluating the density of the points (Section 3.3.1), it can not be used to train a model able to individuate outliers on new traffic traces. Although we cannot perform the same evaluation of results done with K-means, we use it to analyze how the so density-based-constructed training set could affect the semi-supervised AD performances.

Table 3.2: Performances of the different AD approaches evaluated on the *Fiesta de San Cayetano* week (week 6).

	F	P	R
nRNTI - pure K-means	0.0961	1.0	0.0505
RBDowN - pure K-means	0.0329	0.425	0.0171
RBUp - pure K-means	0.0139	0.5	0.0071
nRNTI - LSTM + K-means	0.2535	0.8041	0.1505
RBDowN - LSTM + K-means	0.0296	0.4545	0.0153
RBUp - LSTM + K-means	0.0136	0.5294	0.0069
nRNTI - LSTM + DBSCAN	0.4670	0.8026	0.3293
RBDowN - LSTM + DBSCAN	0.0208	0.4118	0.0107
RBUp - LSTM + DBSCAN	0.0165	0.4583	0.0084

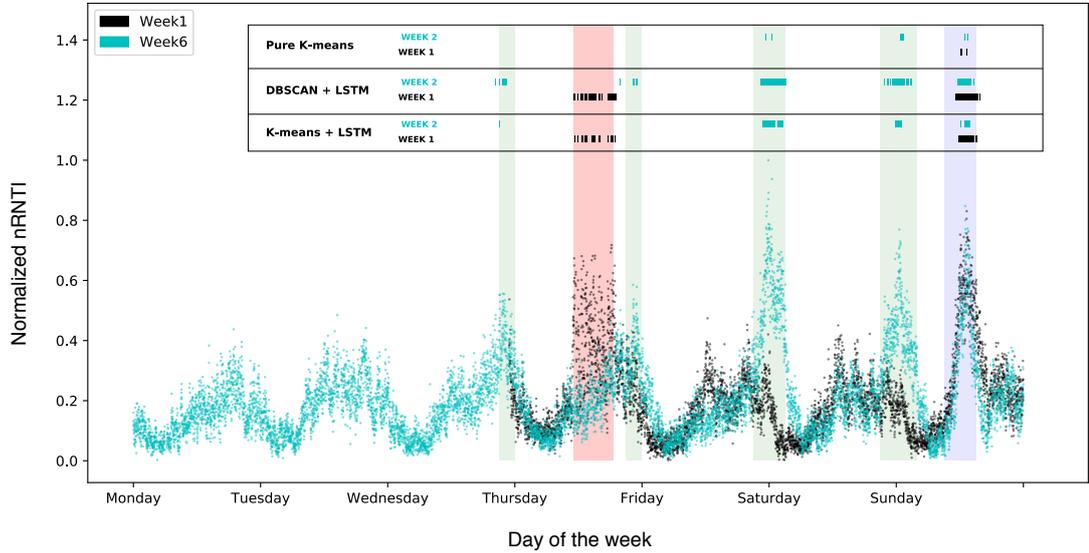


Figure 3.12: Comparison between outliers period identified by K-means and DBSCAN based semi-supervised approach.

3.4.2 Semi-supervised Approach Results Analysis

The LSTM-based AD performances have been evaluated using *Google Colaboratory*, that provides free hardware acceleration with *Tensor Processing Unit (TPU)*. The dataset is split into training and test set so that the test set contains information

about two weeks (the first and the sixth weeks, which are removed from the dataset before the data pre-processing procedure) and all the points tagged as abnormal by the unsupervised methodologies. Moreover the input dataset is divided into training and validation sets with a ratio of 80% - 20%. We implemented the anomaly detection algorithm in *Python*, using *keras* library and *Tensorflow* as backend.

The *Mean Squared Error (MSE)* related to the predictions, both using K-means and DBSCAN methodologies as principle for the training set construction, is similar for all the variables of interest. Although slightly higher performances are related to the DBSCAN-based approach, in general the error is between 0.2-0.25%.

Despite a good prediction represents a positive starting point, for our AD procedure turns out to be of great importance the definition of the p parameter as introduced in 3.3.2. Looking at the points identified as outliers with respect to the variation of p , we fix it for each variable in order to maximize the F value calculated by comparing the obtained anomalies with the abnormal points detected by the pure unsupervised approaches.

As an example, we graphically show in Figure 3.13 the procedure with which the value of p has been defined for the semi-supervised approach based on the DBSCAN methodology. The parameter p represents the percentage of values that the *absolute error* on the prediction, can assume so that the corresponding point is labelled as non-anomalous. For this reason, the vector of the *absolute errors* calculated on the test set is increasingly ordered and p is varied in the range 85%-100%. We confront the vector of the anomalous points detected by the unsupervised techniques with the one obtained by the predictive algorithm when changing the parameter p . Each p value will be matched by different abnormal-defined composition of points. With high p , the number of points defined as outliers will be little, positively influencing the value of P that will grow because these few anomalies will be among those

defined as such also by the unsupervised approach. At the same time R will be negatively influenced because of the inability to identify major parts of what the unsupervised algorithm defines as outliers (Figure 3.14c , Figure 3.14d). Similarly, low p will correspond to high R values and low P values (Figure 3.14a). F represents the perfect mediation between these values and for this reason, at the end, p is fixed equal to the value able to maximize it (Figure 3.14b).

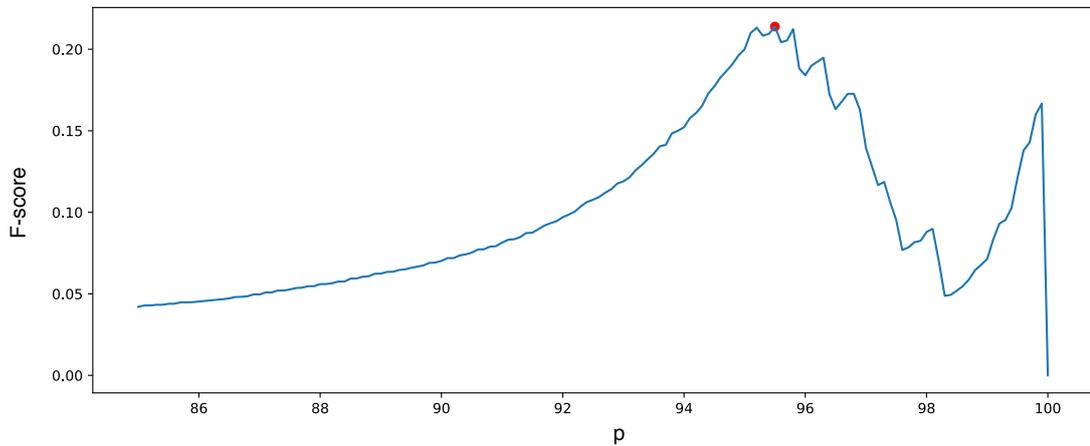


Figure 3.13: p definition for DBSCAN-based semi-supervised approach.

Once p is fixed, the *absolute error* of the prediction is evaluated for each point to define which ones actually represent an anomaly.

The semi-supervised approach applied to the nRNTI metric is able to find contextual anomalies in both the weeks composing the LSTM test set. Even if it was expected during the sixth week, the procedure finds an anomalous pattern also during the Thursday of the first week. Performing a targeted research on this day (30/06), we discover that it is related to the *Orgullo Gay manifestation*. Although a detailed program of the manifestation is not available, we suppose that the contextual anomaly underlined in red in Figure 3.12 could be related to the parade of the pride. This consideration is done looking at the sudden change in

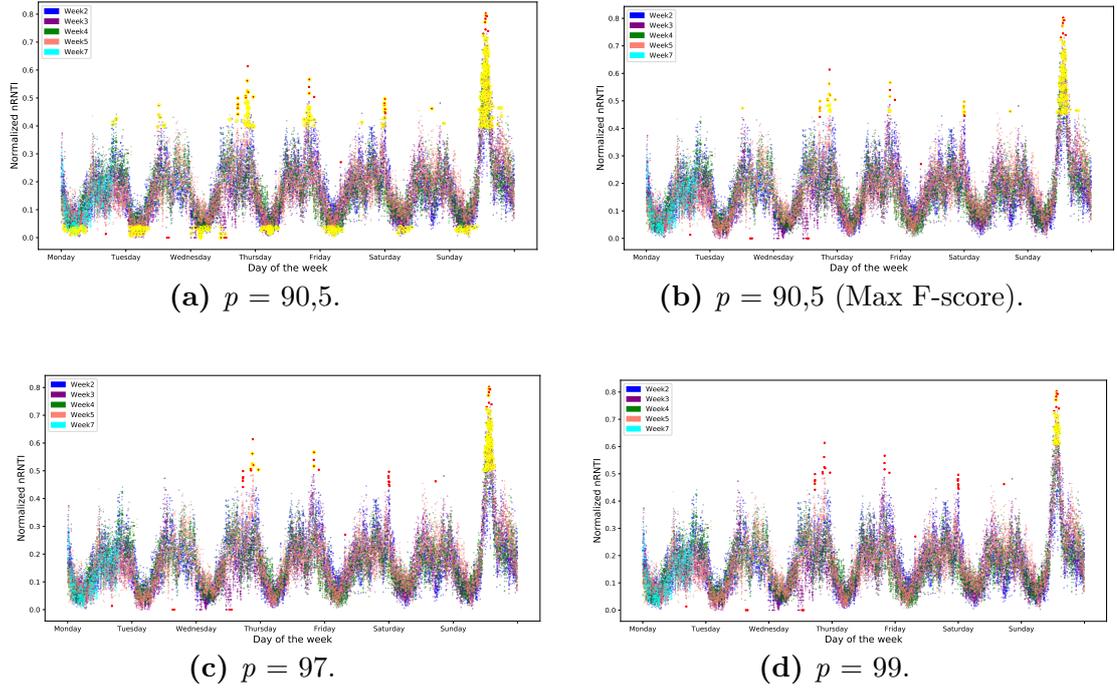


Figure 3.14: Anomalies identified by DBSCAN (red point) and by the semi-supervised LSTM-based procedure (yellow circles) for different values of p .

the nRNTI in the identified anomalous period.

Following this discovery we tested the ability of the pure K-means approach to identify even this type of anomaly. In the wake of the poor results already achieved over the sixth week, it has been found that the mere unsupervised approach is unable to detect it.

As for the pure K-means approach we evaluate the performance in terms of F , P and R using the common test set. Looking at the F-score values in Table 3.2, using DBSCAN as unsupervised basis for the training set construction, is possible to perform a better detection of the relevant time intervals. The precision in identifying anomalous nRNTI turns out to be quite high and the lower R represents the inability of the AD procedure in identifying all the anomalous points, but ensures that those defined as such are truly so. Even if the P metric results to increase a bit using the K-means based approach, the AD procedure tends to

classify as normal many points representing an anomaly, affecting the F-score.

Figure 3.12 reports the contextual anomalies identified using the nRNTI variable. The semi-supervised approach is able to identify anomalies during all the periods of interest related to the *Fiesta de San Cayetano* and it works pretty good in the interval related to the *Rastro Market*, during Sunday. We emphasize its capacity in identifying the beginning of the event a couple of hours after the actual opening: on Sundays people will tend to postpone the normal activity for a few hours, reaching the maximum attendance around lunchtime, as expected from such an event.

The identification of the abnormal events using resDown and resUp variables produce bad results. The calculated F values are very low, using both the unsupervised techniques as the basis to construct the training set. The R values reveal us that this is related to the inability of the AD procedures to detect a good part of the abnormal points. Considering that the events we are trying to detect imply an unusual number of users in the area of interest, this behavior could be supported by the low correlation, shown in Figure 3.3, between the active number of users (nRNTI) and the variables related to the resource blocks (resDown and resUp): even the RNN has not been able to find a hidden relationship between our variables.

3.5 Conclusion

The ability to recognize periods of contextual anomalies is straightly related to the efficient management of cities and it is of great importance for government and public administration, which could benefits even more if anomalies can be automatically alerted in their early stage.

In this chapter, we studied an AD problem using the support of real-world PDCCH dataset collected in the Rastro district, in the city of Madrid. Through the proposed AD framework, we are able to identify anomalous events caused by

crowd gathering by processing the information directly retrievable from the DCI message exchanged between the eNodeB and the connected users over time.

The semi-supervised approach that we develop, enables to identify different type of events without using a-priori information, but rather using unsupervised ML techniques to gain confidence in mobile traffic demand patterns and build an ad-hoc ground truth. In particular, the ability of DBSCAN to identify abnormal points based on their recursivity, combined with the potential to learn the temporal contexts of the input sequences of LSTM networks, results to be a winning combination for urban contextual anomaly detection.

Our evaluation indicates that the proposed framework and the algorithm designed are general enough to identify crowd events in metropolitan areas overcoming the performance of the pure K-means approach with which we perform the comparison.

Chapter 4

Conclusion

The growing pervasiveness of mobile technologies and the enormous amount of data obtainable from them, open new opportunities to monitor and manage the mobile networks elements, each with its own complexity and diversity. In a context in which the traffic demand from users is continuously growing, and the service outages are not admitted for their high societal costs, we decide to exploit the capability of ML solutions in the effort to deepen people's knowledge of habits, and to investigate the future availability of renewable power sources.

That is because the next generation systems have the ability to be dynamically reconfigurable, driven by analytics and intelligence. In this sense, the study of mobile network users' habits, enables automatic procedures that may address a multitude of issues, like capacity planning, outage detection and energy saving; to move toward the optimization of the network functioning, simplifying the maintenance and saving costs without negatively impacting the environment. An example could be the development of mobile networks able to support the traffic demand with sustainable solutions, like cellular BSs directly powered from the RES through an autonomous or hybrid design with other means of renewable or non-renewable energy.

However, the feasibility of such solutions is implicitly related to the ability in forecasting Renewable Energy Sources (RES) availability, such as solar and wind in the short and medium term, and in identifying anomalous situations, which could undermine the functioning of such BSs; due to the unexpected traffic demand and the related exceed in energy consumption.

With this consciousness, during the six months of research carried out at the *Centre Tecnològic Telecomunicacions Catalunya (CTTC)* in Castelldefels, we have evaluated the potentiality of deep learning LSTM networks, proceeding in two directions.

At first, we have designed an evaluation strategy for a correct sizing of LSTM networks using as testing ground the Renewable Energy Sources (RES) forecasting topic; so that the final LSTM's architecture represents a trade-off between the accuracy of the prediction on the test set and the complexity of the network. We employ data available online through the *SoDa HelioClim3* web service, which provides climatic reports from which we extrapolate those related to Castelldefels, with a time granularity of 15 minutes, for a temporal interval of two years.

As show the results in Chapter 2, the constructed LSTM networks perform very well in Wind Speed (*WS*) and Global Horizontal Irradiance (*GHI*) forecasting at one time-step ahead. Indeed, they provide better results with respect to a similar constructed Feedforward Neural Network (FNN).

Then, using the same sizing procedure, we define the LSTM architecture that, exploiting the mobile network as an additional sensing platform to monitor large metropolitan areas, uses the information exchanged over time by the different network elements (e.g., base stations, mobile terminals) for contextual Anomalies Detection (AD) purpose. In particular, we have developed a semi-supervised approach that windows the historical information labeled as normal by two different unsupervised learning techniques (namely, DBSCAN and K-means) to train the

LSTM network. By doing so, when applied to new input traces, it will produce high prediction errors on anomalous data. As a result, the AD problem is not addressed as a supervised classification task, but rather, our algorithm is taught to detect traffic anomalies learning only from non-anomalous examples.

The information used in our approach is directly accessible from the DCI message exchanged between an eNodeB and its connected users in the Rastro district, in the city of Madrid. Looking at the DCI message, at each Transmission Time Interval (TTI = 1 ms), we can access: the number of active users (nRNTI), and the number of allocated resource blocks, both in Uplink (RBU_{up}) and Downlink (RBD_{down}).

Although we demonstrate that the potential of LSTM is not fully exploitable using the details about RBD_{down} and RBU_{up}, knowing the nRNTI connected to the base station with a one-minute granularity, we can identify events related to particular occurrences such as the *Fiesta De San Cayetano* and the *Orgullo Gay* manifestation.

The capacity of our semi-supervised approach far exceeds the mere K-means approach with which it is compared and, more specifically, it results that using DBSCAN as the unsupervised technique for the construction of the training set, it is possible to reach the best performances in terms of detection of anomalous periods. The ability of DBSCAN to identify abnormal points based on their recursivity, combined with the potential to learn the temporal contexts of the input sequences of LSTM networks results to be, from our studies, a winning combination for urban contextual anomaly detection.

Despite the potential of deep learning techniques in the fields of application that we have tested in our studies, the work and the effort to get to a fully sustainable mobile network are still multiple and complex. It is necessary to define modalities of integration between renewable and non-renewable systems, that can lead to global coverage of the cellular network, which includes also isolated and rural areas,

allowing a total sharing of the benefits of new technological progress.

This requires robust systems able to forecast with high accuracy long-term variable RES such as wind, that despite its mutable behavior is less subject to seasonal variations with respect solar, and represent a solution for cloudy and rainy days. In the same way, the detection of contextual anomalies could be done in a more accurate way by employing more efforts in the combination of information provided by different sources, and daily exchanged by the network actors.

What is clear is that today's unprecedented technological advance involves major management problems and environmental impact that cannot and must no longer be ignored, given the amount of possible new solutions still to be investigated.

Bibliography

- [1] *Cisco Visual Networking Index: Forecast and Trends, 2017–2022 White Paper*. <https://www.cisco.com/c/en/us/solutions/collateral/service-provider/visual-networking-index-vni/white-paper-c11-741490.html> (cit. on pp. 10, 11).
- [2] Jingjin Wu, Yujing Zhang, Moshe Zukerman, and Edward Yung. «Energy-Efficient Base-Stations Sleep-Mode Techniques in Green Cellular Networks: A Survey». In: *IEEE Communications Surveys Tutorials* 17 (Apr. 2015), pp. 1–24. DOI: 10.1109/COMST.2015.2403395 (cit. on p. 11).
- [3] Xiang Cheng, Luoyang Fang, Xuemin Hong, and Liuqing Yang. «Exploiting Mobile Big Data: Sources, Features, and Applications». In: *IEEE Network* 31 (Jan. 2017), pp. 72–79. DOI: 10.1109/MNET.2017.1500295NM (cit. on p. 12).
- [4] Kan Zheng, Zhe Yang, Kuan Zhang, Periklis Chatzimisios, Kan Yang, and Wei Xiang. «Big data-driven optimization for mobile networks toward 5G». In: *IEEE Network* 30 (Jan. 2016), pp. 44–51. DOI: 10.1109/MNET.2016.7389830 (cit. on p. 12).
- [5] C. Jiang, H. Zhang, Y. Ren, Z. Han, K. Chen, and L. Hanzo. «Machine Learning Paradigms for Next-Generation Wireless Networks». In: *IEEE Wireless Communications* 24.2 (Apr. 2017), pp. 98–105. ISSN: 1558-0687. DOI: 10.1109/MWC.2016.1500356WC (cit. on p. 12).

- [6] Monica Paolini and Senza Fili. *Mastering Analytics: How to benefit from big data and network complexity* (cit. on p. 12).
- [7] Chaoyun Zhang, Paul Patras, and Hamed Haddadi. «Deep Learning in Mobile and Wireless Networking: A Survey». In: *IEEE Communications Surveys Tutorials* PP (Mar. 2018). DOI: 10.1109/COMST.2019.2904897 (cit. on pp. 13, 17).
- [8] Mohammed H. Alsharif and Jeong Kim. «Optimal Solar Power System for Remote Telecommunication Base Stations: A Case Study Based on the Characteristics of South Korea’s Solar Radiation Exposure». In: *Sustainability* 8.9 (2016). ISSN: 2071-1050. DOI: 10.3390/su8090942. URL: <https://www.mdpi.com/2071-1050/8/9/942> (cit. on p. 14).
- [9] Asma Mohamad Aris and Bahman Shabani. «Sustainable Power Supply Solutions for Off-Grid Base Stations». In: *Energies* 8.10 (2015), pp. 10904–10941. ISSN: 1996-1073. URL: <https://www.mdpi.com/1996-1073/8/10/10904> (cit. on pp. 14, 21).
- [10] Mohammed Alsharif, Jeong Kim, and Jin Kim. «Green and Sustainable Cellular Base Stations: An Overview and Future Research Directions». In: *Energies* 10 (Apr. 2017), p. 587. DOI: 10.3390/en10050587 (cit. on pp. 14, 20, 21).
- [11] André Gensler, Janosch Henze, Bernhard Sick, and Nils Raabe. «Deep Learning for solar power forecasting — An approach using AutoEncoder and LSTM Neural Networks». In: (Oct. 2016), pp. 002858–002865. DOI: 10.1109/SMC.2016.7844673 (cit. on p. 16).
- [12] Rich H. Inman, Hugo T.C. Pedro, and Carlos F.M. Coimbra. «Solar forecasting methods for renewable energy integration». In: *Progress in Energy and Combustion Science* 39.6 (2013), pp. 535–576. ISSN: 0360-1285. DOI:

- <https://doi.org/10.1016/j.pecs.2013.06.002>. URL: <http://www.sciencedirect.com/science/article/pii/S0360128513000294> (cit. on p. 16).
- [13] Gregor Giebel, Richard Brownsword, George Kariniotakis, Michael Denhard, and Caroline Draxl. «The State of the Art in Short-Term Prediction of Wind Power A Literature Overview, 2nd Edition». In: (Jan. 2011). DOI: 10.13140/RG.2.1.2581.4485 (cit. on p. 16).
- [14] Alexandre Costa, Antonio Crespo, Jorge Navarro, Gil Lizcano, Henrik Madsen, and Everaldo Feitosa. «A review on the young history of the wind power short-term prediction». In: *Renewable and Sustainable Energy Reviews* 12.6 (2008), pp. 1725–1744. ISSN: 1364-0321. DOI: <https://doi.org/10.1016/j.rser.2007.01.015>. URL: <http://www.sciencedirect.com/science/article/pii/S1364032107000354> (cit. on p. 16).
- [15] R. L. Welch, S. M. Ruffing, and G. K. Venayagamoorthy. «Comparison of feedforward and feedback neural network architectures for short term wind speed prediction». In: (June 2009), pp. 3335–3340. ISSN: 2161-4407. DOI: 10.1109/IJCNN.2009.5179034 (cit. on p. 16).
- [16] T. G. Barbounis and J. B. Theocharis. «Locally Recurrent Neural Networks for Wind Speed Prediction Using Spatial Correlation». In: *Inf. Sci.* 177.24 (Dec. 2007), pp. 5775–5797. ISSN: 0020-0255. DOI: 10.1016/j.ins.2007.05.024. URL: <https://doi.org/10.1016/j.ins.2007.05.024> (cit. on p. 16).
- [17] Wu Ji and Keong Chan Chee. «Prediction of hourly solar radiation using a novel hybrid model of ARMA and TDNN». In: *Solar Energy* 85.5 (2011), pp. 808–817. ISSN: 0038-092X. DOI: <https://doi.org/10.1016/j.solener.2011.01.013>. URL: <http://www.sciencedirect.com/science/article/pii/S0038092X11000259> (cit. on p. 16).

- [18] Shikhar Srivastava and Stefan Lessmann. «A comparative study of LSTM neural networks in forecasting day-ahead global horizontal irradiance with satellite data». In: *Solar Energy* 162 (Mar. 2018), pp. 232–247. DOI: 10.1016/j.solener.2018.01.005 (cit. on p. 16).
- [19] Rui Fukuoka, Hiroshi Suzuki, Takahiro Kitajima, Akinobu Kuwahara, and Takashi Yasuno. «Wind Speed Prediction Model Using LSTM and 1D-CNN». In: *Journal of Signal Processing* 22 (July 2018), pp. 207–210. DOI: 10.2299/jsp.22.207 (cit. on p. 16).
- [20] Raghavendra Chalapathy and Sanjay Chawla. *Deep Learning for Anomaly Detection: A Survey*. 2019. arXiv: 1901.03407 [cs.LG] (cit. on p. 17).
- [21] C. Feng, S. Arshad, R. Yu, and Y. Liu. «Evaluation and Improvement of Activity Detection Systems with Recurrent Neural Network». In: *2018 IEEE International Conference on Communications (ICC)*. May 2018, pp. 1–6. DOI: 10.1109/ICC.2018.8422896 (cit. on p. 17).
- [22] C. Huang, C. Chiang, and Q. Li. «A study of deep learning networks on mobile traffic forecasting». In: *2017 IEEE 28th Annual International Symposium on Personal, Indoor, and Mobile Radio Communications (PIMRC)*. Oct. 2017, pp. 1–6. DOI: 10.1109/PIMRC.2017.8292737 (cit. on p. 17).
- [23] Imad Alawe, Adlen Ksentini, Yassine Hadjadj-Aoul, and Philippe Bertin. «Improving Traffic Forecasting for 5G Core Network Scalability: A Machine Learning Approach». In: *IEEE Network* 32 (Nov. 2018), pp. 42–49. DOI: 10.1109/MNET.2018.1800104 (cit. on p. 17).
- [24] Longbiao Chen, Dingqi Yang, Daqing Zhang, Cheng Wang, Jonathan Li, and Thi-Mai-Trang Nguyen. «Deep mobile traffic forecast and complementary base station clustering for C-RAN optimization». In: *Journal of Network and Computer Applications* 121 (2018), pp. 59–69. ISSN: 1084-8045. DOI:

- <https://doi.org/10.1016/j.jnca.2018.07.015>. URL: <http://www.sciencedirect.com/science/article/pii/S1084804518302455> (cit. on p. 17).
- [25] Hoang Duy Trinh, Lorenza Giupponi, and Paolo Dini. «Urban Anomaly Detection by processing Mobile Traffic Traces with LSTM Neural Networks». In: *Proceedings of the 2019 IEEE International Conference on Sensing, Communication and Networking(SECON)* (2019) (cit. on p. 17).
- [26] Vaishali Ganganwar. «An overview of classification algorithms for imbalanced datasets». In: *International Journal of Emerging Technology and Advanced Engineering* 2.4 (2012), pp. 42–47 (cit. on p. 18).
- [27] Hatti Mustapha, Mustapha Hatti, doumia, hamid, koucha, and samir. *Artificial Intelligence in Renewable Energetic Systems*. Jan. 2018 (cit. on p. 20).
- [28] Y. Chen, S. Zhang, S. Xu, and G. Y. Li. «Fundamental trade-offs on green wireless networks». In: *IEEE Communications Magazine* 49.6 (June 2011), pp. 30–37. ISSN: 1558-1896. DOI: 10.1109/MCOM.2011.5783982 (cit. on p. 21).
- [29] *SoDa HelioClim3 web-service*. <http://www.soda-pro.com/web-services/radiation/helioclim3-archives-for-free> (cit. on p. 22).
- [30] Diederik P. Kingma and Jimmy Ba. *Adam: A Method for Stochastic Optimization*. 2014. arXiv: 1412.6980 [cs.LG] (cit. on p. 28).
- [31] United Nations, Department of Economics and Social Affairs. *World Population Prospects: The 2018 Revision*. <https://population.un.org/wup/Publications/Files/WUP2018-Report.pdf>. 2018 (cit. on p. 34).
- [32] Huichu Zhang, Yu Zheng, and Yong Yu. «Detecting urban anomalies using multiple Spatio-temporal data sources». In: *Proceedings of the ACM on*

- Interactive, Mobile, Wearable and Ubiquitous Technologies* 2.1 (2018), p. 54 (cit. on p. 35).
- [33] Afif Osseiran et al. «Scenarios for 5G mobile and wireless communications: the vision of the METIS project». In: *IEEE Communications Magazine* 52.5 (2014), pp. 26–35 (cit. on p. 35).
- [34] Hoang Duy Trinh, Engin Zeydan, L. Giupponi, and Paolo Dini. «Detecting Mobile Traffic Anomalies through Physical Control Channel Fingerprinting: a Deep Semi-supervised Approach». In: *IEEE Access* PP (Oct. 2019), pp. 1–1. DOI: 10.1109/ACCESS.2019.2947742 (cit. on p. 36).
- [35] Victoria Hodge. «A Survey of Outlier Detection Methodologies». In: *Artificial Intelligence Review* 22 (Oct. 2004), pp. 85–126. DOI: 10.1023/B:AIRE.0000045502.10941.a9 (cit. on p. 38).
- [36] Trupti Kodinariya and P.R. Makwana. «Review on Determining of Cluster in K-means Clustering». In: *International Journal of Advance Research in Computer Science and Management Studies* 1 (Jan. 2013), pp. 90–95 (cit. on p. 41).
- [37] Martin Ester, Hans-Peter Kriegel, Jörg Sander, and Xiaowei Xu. «A Density-Based Algorithm for Discovering Clusters in Large Spatial Databases with Noise». In: *KDD*. 1996, pp. 226–231. URL: <http://www.aaai.org/Library/KDD/1996/kdd96-037.php> (cit. on p. 43).
- [38] *Fiesta De San Cayetan official program, 2016*. <https://www.madrid.es/UnidadesDescentralizadas/UDCMedios/noticias/2016/07Julio/29Viernes/Notasprensa/Fiestas%20Centro/ficheros/Programa%20fiestas%202016.pdf> (cit. on p. 47).