future internet

Article

A GIS-Based Hot and Cold Spots Detection Method by Extracting Emotions from Social Streams

Barbara Cardone [1], Ferdinando Di Martino [1,2,*] and Vittorio Miraglia [1]

[1] Dipartimento di Architettura, Università degli Studi di Napoli Federico II, via Toledo 402, 80134 Napoli, Italy
[2] Centro di Ricerca Interdipartimentale "Alberto Calza Bini", Università degli Studi di Napoli Federico II, via Toledo 402, 80134 Napoli, Italy
* Correspondence: fdimarti@unina.it; Tel.: +39-081-253-8907; Fax: +39-081-253-8905

Abstract: Hot and cold spot identification is a spatial analysis technique used in various issues to identify regions where a specific phenomenon is either strongly or poorly concentrated or sensed. Many hot/cold spot detection techniques are proposed in literature; clustering methods are generally applied in order to extract hot and cold spots as polygons on the maps; the more precise the determination of the area of the hot (cold) spots, the greater the computational complexity of the clustering algorithm. Furthermore, these methods do not take into account the hidden information provided by users through social networks, which is significant for detecting the presence of hot/cold spots based on the emotional reactions of citizens. To overcome these critical points, we propose a GIS-based hot and cold spot detection framework encapsulating a classification model of emotion categories of documents extracted from social streams connected to the investigated phenomenon is implemented. The study area is split into subzones; residents' postings during a predetermined time period are retrieved and analyzed for each subzone. The proposed model measures for each subzone the prevalence of pleasant and unpleasant emotional categories in different time frames; with the aid of a fuzzy-based emotion classification approach, subzones in which unpleasant/pleasant emotions prevail over the analyzed time period are labeled as hot/cold spots. A strength of the proposed framework is to significantly reduce the CPU time of cluster-based hot and cold spot detection methods as it does not require detecting the exact geometric shape of the spot. Our framework was tested to detect hot and cold spots related to citizens' discomfort due to heatwaves in the study area made up of the municipalities of the northeastern area of the province of Naples (Italy). The results show that the hot spots, where the greatest discomfort is felt, correspond to areas with a high population/building density. On the contrary, cold spots cover urban areas having a lower population density.

Keywords: GIS; hot spots; cold spots; fuzzy partition; emotional categories; social streams

Citation: Cardone, B.; Di Martino, F.; Miraglia, V. A GIS-Based Hot and Cold Spots Detection Method by Extracting Emotions from Social Streams. *Future Internet* **2023**, *15*, 23. https://doi.org/10.3390/fi15010023

Academic Editors: Dino Giuli and Filipe Portela

Received: 28 November 2022
Revised: 21 December 2022
Accepted: 29 December 2022
Published: 30 December 2022

1. Introduction

Detection and classification of emotions from social streams are one of the main goals of Sentiment Analysis (SA). Generally, Deep Learning algorithms are executed to detect emotions from massive data [1]. However, in cases where the size of the available labeled data is small, it is difficult and expensive to construct large training sets; using semi-supervised and unsupervised algorithms that also employ unlabeled data is an effective option [2].

Semi-supervised learning algorithms [3] use large unlabeled data in combination with the available labeled data in various learning processes. They can improve classification performances using unlabeled data to extract relevant information necessary for classifying data points [4].

Unsupervised learning methods are applied for mining data when no annotation is available; various clustering algorithms were used to detect embedded knowledge in massive data streams [5].

In [6], a lightweight model based on an extension of the Fuzzy C-means algorithm (FCM) [7,8], called Extended Fuzzy C-Means (EFCM) [9], is applied to social streams for emotion-based classification. Using this procedure, documents are classified into emotional categories by being assigned to the emotional category that corresponds to the fuzzy cluster to which it has the highest membership degree. The authors showed that this approach produces good classification accuracy.

This model is implemented in a Geographical Information System (GIS) in [10] to classify urban districts based on the orientation of citizens detected from social streams.

The document classification model [6] represents a significant trade-off between classification accuracy and computational complexity. However, it has two main critical points:

- The number of clusters must match the number of emotional categories, and there can be no a priori one-to-one mapping between clusters and emotional categories;
- The centers of the initial clusters are set randomly. This produces an average increase in execution times and allows the algorithm to converge toward a local minimum;
- the document is assigned to a single emotional category, not considering emotional categories corresponding to clusters to which the document belongs with a non-negligible membership degree.

A new framework for emotion-based classification from social streams, using an entropy-based weighted variant of FCM, named EwFCM, to classify the data gathered from streams, is provided to overcome the cluster center's initialization problem in [11]. EwFCM applies a fuzzy entropy measure to optimize the initialization of the cluster centers. Test results applied on Twitter datasets show that this emotion classification framework provides better results than the framework proposed in [10] in terms of classification accuracy and execution time. However, this approach does not solve the problems connected to the constraint of fixing the number of clusters equal to the number of emotional categories and to the choice of neglecting the clusters to which the document belongs with a secondary but not negligible membership degree.

To overcome these drawbacks, a new fuzzy emotion-based classification method, called FREDoC (Fuzzy Relevance Emotions Document Classification), is proposed in [12] in which, instead of executing a fuzzy clustering approach to classify a document by assigning it to the prevailing emotional category, an emotion relevance index is employed to determine if an emotional category in a text is relevant, and it is based on the well-known TF-IDF metric. Fuzzy linguistic labels are used to explain the emotional relevance, defined by constructing a specific fuzzy partition; the documents are classified considering those emotional categories whose relevance in the document is not negligible.

In [13], the FREDoC emotion-based multiclassification method is implemented in a GIS framework to classify urban areas by the more relevant critical issues highlighted by citizens.

One of the main problems faced today in spatial analysis is the detection of urban or territorial areas where a specific phenomenon refers to hot spot analysis.

Formally, areas of the study area that have a high/low density of occurrences caused by the phenomenon under consideration are classified as hot/cold patches.

Statistical indices, such as Getis-Ord Gi* statistic [14] and Local Index of Spatial Autocorrelation (LISA) [15], are used to detect hot and cold spots; Gi* evaluates the spatial association of a spatial entity with surrounded ones within a specified distance of a single point. LISA statistics detect spatial autocorrelation between neighboring features, measuring the degree of similarity of data points close to a given data point within a specified threshold distance. LISA is used in [16] for drought hot spot detection in northern Bangladesh.

Kernel Density Estimation (KDE) [17,18] is another well-known hot and cold spot detection technique; by identifying hot/cold spots as zones with high/low-intensity peaks relative to the surrounding areas, it provides a continuous surface representing the geographical distribution of the intensity of the phenomena.

These methods, while distinguishing hot and cold spots from areas in which the intensity of the phenomenon appears normal, are not able to accurately detect the location and extent of hot and cold spots.

To overcome this drawback, some authors propose partitive cluster algorithms to detect the extension and the shape of hot and cold spots. K-means [19] is implemented in GIS environments to detect hot and cold spots in crime [20–22] and fire analysis [23,24].

Recently, K-medoids [25] was applied to detect hot spots in crime [26] and disease analysis [27], and FCM [8,28] was used to detect hot spots in crime analysis [29,30] and road traffic crashes [31].

One of the main defects of hot spot detection methods based on partitive clustering algorithms is the impossibility of detecting the shape of hot and cold spots. In order to accurately detect the shape of the hot spots in [32], a method of hot spot detection was adopted which makes use of the density-based Fast DBSCAN clustering algorithm [33]; in [34], a hybrid clustering combining a self-organizing map and a hierarchical clustering method is applied in a GIS platform to detect service levels of urban streets from GPS flow speeds data points.

Recently EFCM was implemented in a GIS framework in [35] to detect hot spots in fire analysis and in [36] to detect the spatial distribution and the evolution of hot spots in disease analysis. The shape of a hotspot is approximated by a circle on the map. This approach represents a trade-off between the computational speed of hot and cold spot detection methods based on partitive clustering methods and the accuracy of approaches using density-based clustering algorithms in identifying the form of hot and cold areas.

These hot and cold spot detection methods are performed on data points located on the map detecting those areas with a high and low density and intensity of data points. In this paper, we propose a different hot and cold spot detection approach in which areas detected as hot/cold spots are those in which a feeling of unease and discouragement expressed by citizens on social media prevails (it is not very present) and persists over time.

Social data can provide an extensive source of information useful for hot and cold spot analysis, both because they refer to the perceptions and moods of the citizens involved in the phenomenon analyzed and because they represent a massive data source that allows you to analyze how a phenomenon spreads and how it evolves. In addition, in many spatial analysis problems, it is useful to aggregate the information by sub-zone in which the study area is partitioned in order to classify these sub-zones on the basis of the impact of the phenomenon analyzed. The GIS-based framework in [10] responds to this need by classifying these sub-zones on the basis of the perception and moods of citizens detected in posts inserted in social networks.

In many problems, the user is not interested in detecting the exact shape of hot/cold spots on the map but intends to evaluate if a sub-zone of the study area can be classified as a hot or cold spot, that is, if the phenomenon is particularly present or rare on the sub-zone for a defined period of time. This goal was pursued in [35], which investigates whether urban agglomerations the City of London is divided into are more common and which sorts of catastrophic catastrophes endure over time. The idea on which this research is based consists of constructing a GIS-based framework for detecting hot and cold spots based on the information extracted from social streams, with the aim of determining which subzones of the study area are classified as hot/cold spots.

Unlike the method proposed in [35], we do not use fuzzy-clustering-based hot spot detection methods, but we apply an emotion-based classification model to classify sub-zones of the study area based on the relevance of pleasant and unpleasant emotions of citizens determined by starting from social data. This approach allows for taking into account the hidden knowledge extracted from social streams that allow the evaluation of the persistence of the phenomenon in a sub-zone based on the moods of citizens.

Following [6,11,12], we partition the emotions into pleasant and unpleasant, identifying as hot/cold spots a sub-zone classified as unpleasant/pleasant for a duration greater than a predetermined period.

We call our framework FESC (Fuzzy Emotion-based hot and cold Spots Classification).

The study area is partitioned into subzones, and the user defines the atomic time interval, called time frame, to be considered for the analysis of the evolution of the intensity of a phenomenon. A dictionary of emotion categories is created, divided into pleasant and unpleasant. The social data are extracted and grouped for subzone and time frame; after a text parting phase, they constitute the documents of a corpus. The dictionary of emotion categories is structured as in [6], in which each emotion category is assigned a set of terms. The Term Frequency measures the relevance of a term in a document—The Term Frequency—Inverse Document Frequency index (TF-IDF) [37], a well-known metric measuring the importance/relevance of a word in a text, is used to determine the relevance of a term in a document. Following [6,11,12], the TF-IDF values measured for all terms assigned to an emotion category are summed and normalized to form the relevance of this emotion category in the document; then, the TDF-IDF values assigned to pleasant (unpleasant) categories are summed, obtaining a normalized index called pleasant (unpleasant) emotion relevance (for short Pleasant/Unpleasant ER) representing the relevance of pleasant (unpleasant) emotions in the document, where the sum of Pleasant and Unpleasant ER measures in a document is equal to 1.

To evaluate the intensity of the Pleasant/Unpleasant ER in a document, we define a fuzzy partition, built by starting from the emotional Pleasant/Unpleasant ER measure, described by fuzzy linguistic labels.

Each document is classified according to the intensity of the relevance of both pleasant and unpleasant emotions. The linguistic label value of the fuzzy set of the ER fuzzy partition, to which it has the highest membership degree, is used to classify it.

For instance, consider the Low, Medium-low, Medium, Medium-high, and High fuzzy sets as the Pleasant/Unpleasant ER fuzzy partition in Figure 1. Let the pleasant ER measure of a document be equal to 0.59 and the unpleasant ER measure is given by 1 − 0.59 = 0.41.

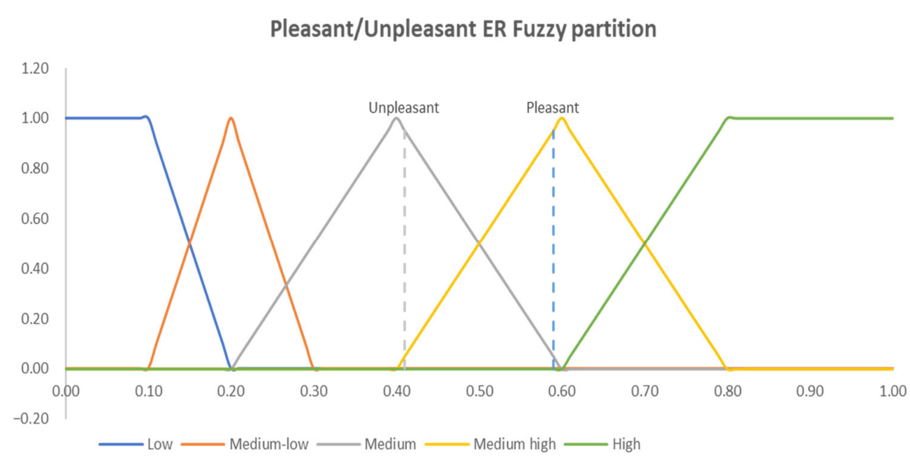


Figure 1. Example of Pleasant/Unpleasant ER fuzzy partition.

Then, the document will be classified into the pleasant class Medium-high and the unpleasant class Medium.

If all the documents corresponding to a subzone's time frames are classified with pleasant ER linguistic labels over a prefixed threshold, this subzone is labeled as a hot spot. Likewise, if those documents are classified with unpleasant ER linguistic labels over that threshold, the subzone is labeled as a cold spot.

Finally, a map of the hot and cold spots of the study area is built.

Figure 2 shows a schematized architecture of FESC.

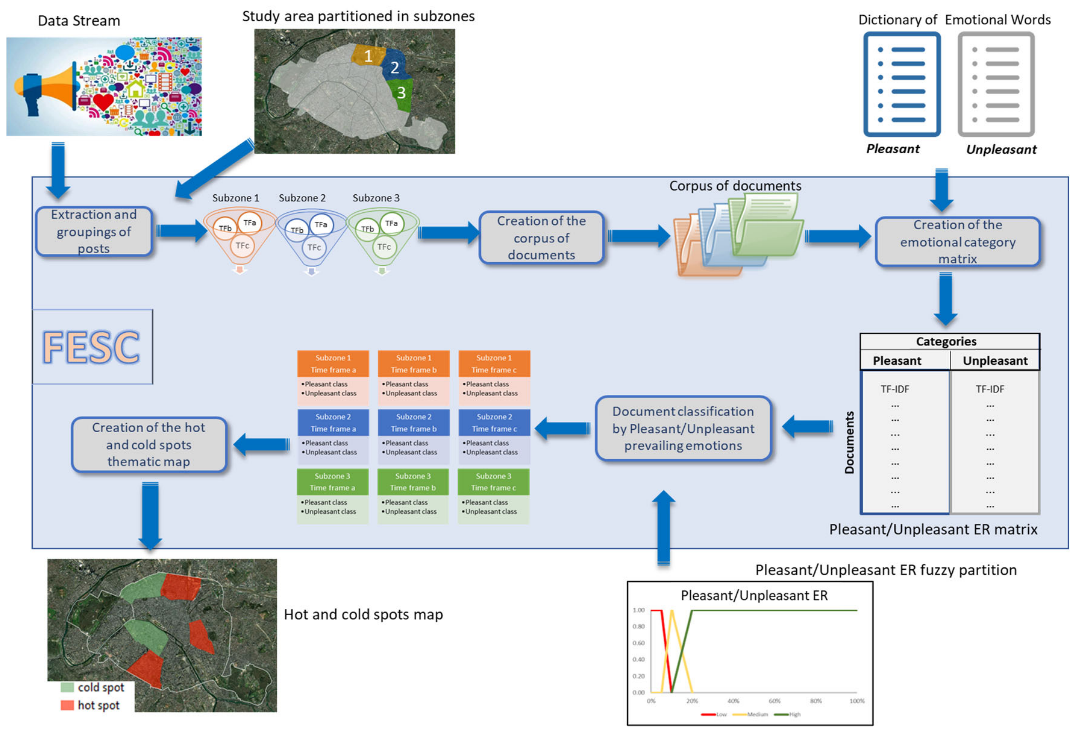


Figure 2. Schema of the proposed GIS-based framework.

Inputs for the framework are the study area divided into subzones, the data stream with the posts entered by citizens in the analyzed period, the Dictionary of the Emotional Word containing the dictionary of terms relating to the single emotional categories, and the ER fuzzy partition.

The output of the process is the hot and cold spots map which highlights the subzones classified as hot and cold spots, corresponding to the zones with unpleasant/pleasant emotions relevance classified above the threshold for all time frames.

A detailed description of the components of the framework is shown in Section 3.

The main benefits of our method can be summarized as follows:

- Unlike other hot and cold spot detection methods, which take into consideration only localized events, FESC recognizes hot and cold spots by analyzing the moods of citizens extracted from social networks; this makes it possible to use hidden knowledge to evaluate, on the basis of citizens' emotions, which subzones of the study area are more critical and which, on the other hand, are less affected by the phenomenon being investigated;
- FESC is computationally faster than cluster-based hot and cold spot detection methods; in fact, FESC does not need to detect the exact geometric shape of a spot but evaluates the prevailing emotional states of citizens to determine which subzones can be classified as hot or cold spots; this mode significantly reduces the CPU times of the algorithm;
- to overcome the critical issues of the cluster-based approaches adopted for the classification of documents based on the prevailing emotional category, which needs to

fix the number of clusters equal to the number of emotional categories and to establish a one-to-one correspondence between clusters and emotional categories, FESC adopts an approach based on the construction of a fuzzy partition of the relevance of pleasant and unpleasant emotional categories.

The remaining portions of the essay are structured as follows: The introduction to formal concepts and ideas on hot (or cold) spots are provided in Section 2, along with a synthetic description of the emotion-based document categorization approach [6]. The logical overview described in Section 3 serves as an introduction to the suggested framework. The findings of the tests and case studies that demonstrate the viability of the suggested strategy are presented in Section 4. We discuss the paper's closing findings in Section 5.

2. Preliminaries

2.1. Hot and Cold Spots in Spatial Analysis

A hot or cold spot in spatial analysis is an area in the research area where a high/low concentration of events defining a particular phenomenon is focused [38].

A technique used to identify these areas is the detection of hot/cold spots. Each event is geometrically represented as a point on the map when a clustering algorithm is used to detect hot and cool regions; the set of events forms the input data points for the clustering algorithm; the resultant cluster prototypes are the hot/cold spots painted as polygons on the map.

For the benefit of the algorithm's faster calculation, it is not always required to identify the precise geometry of hot spots in spatial analysis issues. For example, in [19,28], cluster prototypes are represented by points on the map, and hot spots are made up of areas containing these points. In [34,36], EFCM is executed to detect hot spots as circular areas.

Additionally, to examine how the phenomenon has changed through time, it is necessary to determine how it evolves and where it persists over a period of time. In [36], the dataset is partitioned into subsets; each subset contains data points corresponding to the events that occurred in a given period. To evaluate the temporal evolution of the hotspots that identify disease strains, EFCM is run for each subgroup. In [35], a hot spot detection method based on a fuzzy entropy-based variation of EFCM is executed to analyze the temporal evolution of the intensity of criminal events in each urban agglomeration of the City of London. An urban agglomeration whose extension is significantly covered by clusters detected in several consecutive time frames represents a hot spot for the criminal phenomenon analyzed.

The definition of a hot or a cold spot can be generalized to an area in which a high or low frequency of events is detected for the period of time analyzed.

Let the study area be partitioned into n atomic subzones $s_1, s_2, \ldots, s_n$. Suppose we want to analyze the evolution of the intensity of a phenomenon in a period of time broken down into T time frames having equal duration. The ith subzone can be labeled as a hot/cold spot if the frequency of events occurring in ith is greater or less than a specified threshold.

2.2. Lightwise to Classify Social Messages Emotion Classification Methods

In [6], a framework was proposed to classify documents extracted from social streams based on the prevailing emotional category. An architectural schematization of this framework is shown in Figure 3.

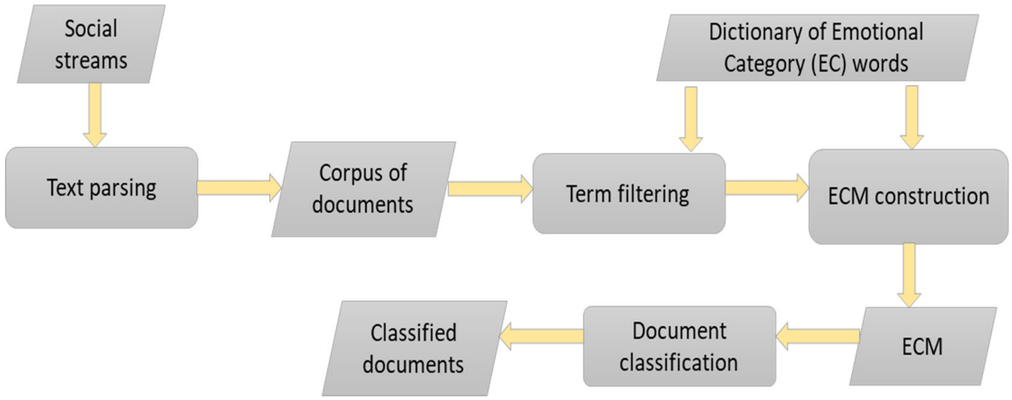


Figure 3. Schema of the social messages emotion classification framework proposed in [6].

Social messages containing specific words are extracted and analyzed by the Text Parsing component; Text parsing filters out irrelevant terms in the texts, executes the Part of Speech tagging, and aggregates the messages having common features such as a common hashtag, constructing the Corpus of Documents.

The Term Filtering component selects from the text in each document all the words expressing emotions. It annotates words related to terms appearing in the Dictionary of Emotional Category Words, a dictionary containing, for each pleasant and unpleasant emotional category, a list of terms corresponding to that emotional category, where each term is reduced to its stemmed form. When Term Filtering annotates a word in the document, it assigns the word to the corresponding term in the Dictionary of Emotional Category Words.

The ECM Construction component creates the Emotional Category Matrix (ECM), a matrix whose elements are given by the TF-IDF measures relating to the relevance of an emotional category in a document.

Formally, let D = $\{d_1, d_2, \ldots d_N\}$ be the Corpus of Documents and t be a term in the Dictionary of Emotional Category words, the TF-IDF index measuring the relevance of the term t in the ith document TF-IDF (t, d_i) is given by the product of two terms:

$$\text{TF-IDF}\ (\mathrm{t}, \mathrm{d_i}) = \mathrm{tf}(\mathrm{t}, \mathrm{d_i}) \cdot \mathrm{idf}(\mathrm{t}, \mathrm{D}) \tag{1}$$

The first term, called Term Frequency, measures the frequency of the term t in document d_i. The last term, called inverse document frequency, measures, in decimal logarithmic scale, the relevance of the term in the document collection.

If f(t, d_i) is the number of times the term t appears in the document d_i and n_i is the number of times all terms appear in the document d_i, the term frequency is given by:

$$\mathrm{tf}(\mathrm{t}, \mathrm{d_i}) = \frac{\mathrm{f}(\mathrm{t}, \mathrm{d_i})}{\mathrm{n_i}} \tag{2}$$

If N_t is the number of documents in which the term t appears at least once, the inverse document frequency is given by:

$$idf(t, D) = \log_{10} \frac{N}{N_t} \tag{3}$$

Let C = $\{c_1, c_2, \ldots c_M\}$ be the set of emotional categories. If T_j is the set of terms related to the emotional category c_j, the relevance of the jth category in the ith document is given by the formula:

$$\text{TF-IDF}\ (\mathrm{c_j}, \mathrm{d_i}) = \sum_{\mathrm{t} \in \mathrm{T_j}} \text{TF-IDF}\ (\mathrm{t}, \mathrm{d_i}) \tag{4}$$

The measure TF-IDF (c_j, d_i) is the component (i, j) of the EC matrix.

The Document Classification component analyzes the EC matrix and classifies each document assigning it to the most relevant emotional category. The EFCM is used to carry out this procedure in [6], with the restriction that the number of clusters equals the number of emotional categories. A mapping that associates an emotional category to each fuzzy cluster is constructed. The document is then labeled with the name of the appropriate emotional category and placed in the cluster to which it has the highest membership degree.

The EwFCM technique, which employs the fuzzy entropy measure to optimize the initialization of cluster centers, is used to replace EFCM in order to improve the clustering performances in [11].

In [12], the authors adopt a different document classification approach; to overcome the problem of the constraint on the number of clusters, instead of using a clustering algorithm to classify documents, they normalize the TF-IDF measures in [0,1] and use a fuzzy partition of the relevance of an emotional category in a document to facilitate the interpretation of the emotional relevance by the user and allow the classification of the document, also considering emotional categories with relevance in the secondary document but not negligible An emotional relevance index is used in [12] to determine the relevance of a category is inside a document. The ratio indicates the emotional relevance of the jth emotional category in the ith document:

$$R(c_j, d_i) = \frac{\text{TF-IDF }(c_j, d_i)}{\sum_{k=1}^{M} \text{TF-IDF }(c_k, d_i)} \tag{5}$$

The value $R(c_j, d_i)$ varies within the interval [0,1]. The relevance assigned to the category c_j in the document d_i is given by the label of the fuzzy set of the category relevance fuzzy partition having the greatest membership degree. When a category's relevance in the document surpasses a certain level, the document is categorized by being assigned to that category.

3. The FESC Framework

In [10], the social messages emotion classification method [6] shown in Figure 3 is implemented in a GIS-based framework in order to classify the districts of the municipality of Bologna (Italy). The study area is given by the municipality, and its six districts are the subzones in which the study area is spatially divided. Social messages are extracted in order to detect the moods of citizens connected to the livability of a district. Each document is related to Twitter social messages posted by citizens and tourists whose hashtags contain the name of a district. Each district is categorized using EFCM according to the dominant emotion category. The districts are then organized thematically depending on the dominant emotion type.

FESC extends this framework to study the temporal evolution of the intensity of a phenomenon in a subzone and detect if a subzone constitutes a hot or cold spot.

Figure 4 shows the logical view of our framework.

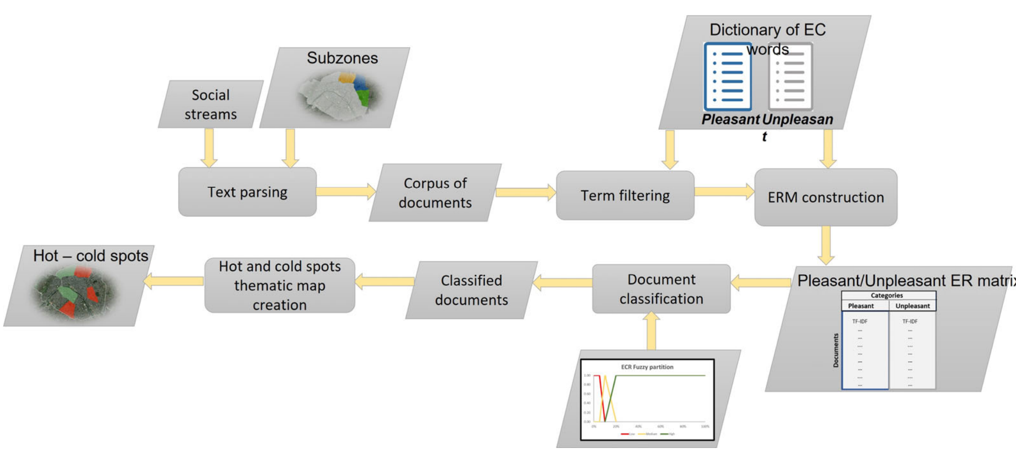


Figure 4. FESC: logical overview.

The social data extracted contain all the messages related to the analyzed phenomenon posted by users located in one of the subzones in which the study area is partitioned and in a specific period. This period is broken down into T time frames having equal duration.

Let the study area be partitioned into n atomic subzones s_1, s_2,…, s_n, and the analyzed period be partitioned in T time frames. After cleaning up the text from irrelevant terms and executing the Part of Speech tagging, the Text Parsing component creates N documents aggregating messages related to the same subzone and the same time frame, where N = n·T. The Term Filtering component annotates terms belonging to pleasant and unpleasant emotional categories in the Dictionary of EC words. The ERM Construction component computes the IF-TDF value of each category in the document and aggregates the values assigned to each pleasant/unpleasant category.

Let $\{c_{p1,}c_{p2}, \dots, c_{pM_p}\}$ be the set of M_p pleasant emotional categories and $\{c_{u1,}c_{u2}, \dots, c_{uM_u}\}$ be the set of M_u unpleasant emotional categories, where $M_p + M_u = M$.

If $R(c_j, d_i)$ j = 1,…,M i = 1,…, N is the relevance of the jth category in the ith document measured by (5), the relevance of the pleasant and unpleasant emotions in the document are given by:

$$R_{p,i} = \sum_{k=1}^{M_p} R(c_{kp}, d_i) \tag{6}$$

$$R_{u,i} = \sum_{k=1}^{M_u} R(c_{ku}, d_i) \tag{7}$$

where $R_{p,i} + R_{u,i} = 1 \; \forall \, i = 1, \dots, N$.

The output of this process is a two-columns matrix called the *Pleasant/Unpleasant ER matrix* in which are assigned, for each document, the relevance of pleasant and unpleasant emotional categories given, respectively, by Equations (6) and (7).

The Document Classification component fuzzifies pleasant and unpleasant category relevancies using a user-built fuzzy pleasant/unpleasant category relevance partition called ECR fuzzy partition. The ith document is classified with pleasant/unpleasant emotion relevance equal to the label of the fuzzy set to which R_{pi} (R_{ui}) belongs with the highest level of membership. The hot and cold spot thematic map creation component analyzes all the documents associated with a subzone, where each document refers to the emotional manifestations and discomforts expressed in social posts residing in the subzone in a given time frame. If all the documents are classified with the relevance of unpleasant emotions above a specified threshold, then the subzone is labeled as a hot spot; likewise, if all documents are classified with the relevance of pleasant emotions above a specified threshold $\hat{R}$, the subzone is labeled as a cold spot.

As an example, we now consider the fuzzy partition in Figure 1. Suppose we set $\hat{R}$ = Medium high as the threshold value for the relevance.

Let d_{h1}, d_{h2},…, d_{hT} be the T documents related to the hth subzone, where d_{ht} is the document referring to the posts inserted in the tth time frame. This subzone will be labeled as a hot spot if all its documents are classified with the relevance of unpleasant emotions greater or equal to Medium high. Likewise, it will be labeled as a cold spot if all its documents are classified with the relevance of pleasant emotions greater or equal to Medium-High.

The FESC algorithm, structured in pseudocode, is shown below.

In the preprocessing phase, the partitioning of the study area in subzones is performed; the analyzed period is partitioned in T time frames, and the pleasant and unpleasant ER fuzzy partition is created. After classifying each document, the number of related documents classified with pleasant/unpleasant emotion relevance greater than or equal to the threshold value $\hat{R}$ is calculated for each subzone. If this number is equal to the number of time frames T, then the subzone is labeled as a cold or hot spot. Finally, the hot and cold spots thematic map is created.

In Algorithm 1 the FESC method is schematized in pseudocode.

Algorithm 1: Fuzzy Emotion-based hot and cold Spots Classification (FESC)

1. Partition the study area in n subzones;
2. Partition the analyzed period in T time frames having an equal time width;
3. Create the pleasant and unpleasant ER fuzzy partition;
4. Create the Dictionary of EC words;
5. Set the relevance threshold $\hat{R}$;
6. Extract the social messages;
7. Clean the irrelevant texts in the posts;
8. Create the Corpus of Documents aggregating the posts by subzone and time frames;
9. **For each** document:
10. Annotate terms belonging to pleasant and unpleasant emotional categories;
11. **For each** emotional category:
12. Compute the IF-TDF value by (4);
13. Calculate the emotional relevance by (5);
14. **Next:**
15. Aggregate the emotional relevance of pleasant emotional categories by (6);
16. Aggregate the emotional relevance of unpleasant emotional categories by (7);
17. Classify the document assigning the pleasant and unpleasant emotion relevances;
18. **Next:**
19. **For** each subzone:
20. n_{cold}: = 0 // number of documents with pleasant emotion relevance ≥ $\hat{R}$;
21. n_{hot}: = 0 // number of documents with unpleasant emotion relevance ≥ $\hat{R}$;
22. **For** each time frame:
23. R_p: = pleasant emotion relevance of the corresponding document;
24. R_u: = unpleasant emotion relevance of the corresponding document;
25. **If** R_p = ≥ $\hat{R}$ **Then:**
26. N_{cold}: = n_{cold} + 1;
27. **End if:**
28. **If** R_u = ≥ $\hat{R}$ **Then:**
29. n_{hot}: = n_{hot} + 1;
30. **End if:**
31. **Next:**
32. **If** n_{cold} = T, **Then:**
33. Label the subzone as a cold spot;
34. **Else:**
35. **If** n_{cold} = T, **Then:**
36. Label the subzone as a hot spot;
37. **End if:**
38. **End if:**
39. **Next:**
40. Create the hot and cold spots thematic map.

---

4. Test and Results

FESC was tested on a study area given by the 18 municipalities of the northeastern area of the province of Naples (Italy). The study area is a peripheral urban area characterized by the presence of a high population and building density. The phenomenon that has been analyzed concerns the discomfort of citizens caused by heat waves that occurred in the study area in the summer seasons between 2020 and 2022. The choice of this study area is dictated by the fact that it is a densely populated suburban area with recent popular building development, which can aggravate the discomfort of citizens during the presence of heat wave phenomena.

The map in Figure 5 shows the study area.

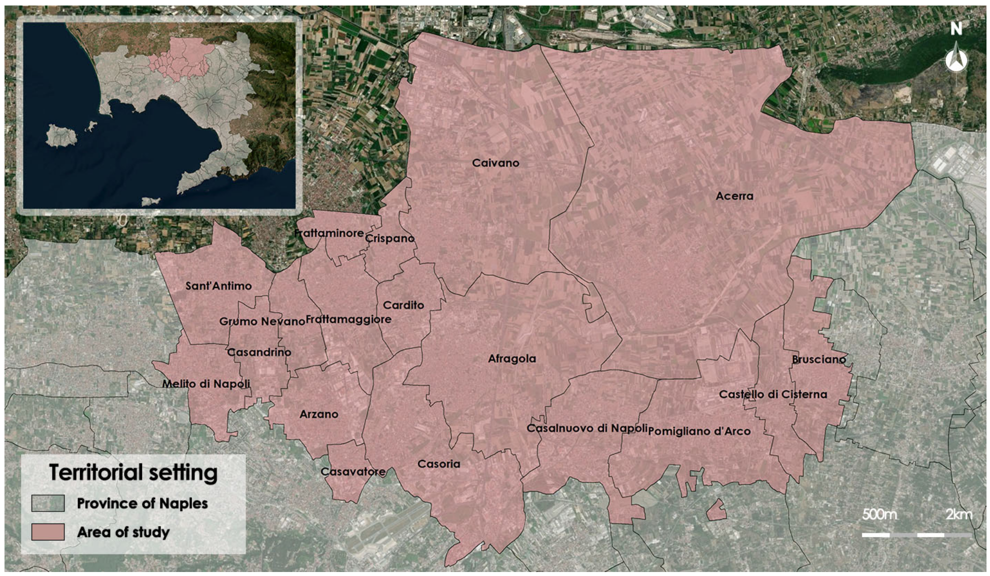


Figure 5. The study area—the 18 municipalities of the northeastern area of the province of Naples (Italy).

To carry out the tests, tweets posted by users residing in the municipalities of the study area were extracted during the summer seasons of the time period analyzed. Each municipality constitutes a subzone, and the time period has been divided into three-time frames corresponding to the years 2020, 2021, and 2022.

The posts were extracted by selecting tweets with hashtags and keywords given at least one of the following words: heat, sultriness, heatwave, warm, scorching heat, and torrid heat.

The Python TWitter INTelligence library (TWINT) was used to extract the selected Twitter posts.

The FESC framework was implemented on a GIS platform built on the ESRI ArcGIS pro suite.

The number of tweets is about 622,100; on average, 2,900 monthly tweets for each municipality. The Python Geocoder library was used to match the tweets to the corresponding municipality. The tests were executed using an Intel Core i7 Pentium having a 2.90 GHz clock frequency. The CPU time to execute the FESC algorithm on this dataset and to detect the final hot and cold spots was about 10.5 min.

After deleting noisy posts, posts referring to the same municipality and time frame have been merged into one document.

To construct the Dictionary of EC words was considered the set of 16 emotional categories, divided into eight pleasant and eight unpleasant, used in [12]; these emotional categories are given by the basic and secondary emotional categories in Plutchik's wheel of emotions [39], listed in Table 1.

Table 1. Pleasant and unpleasant emotional categories.

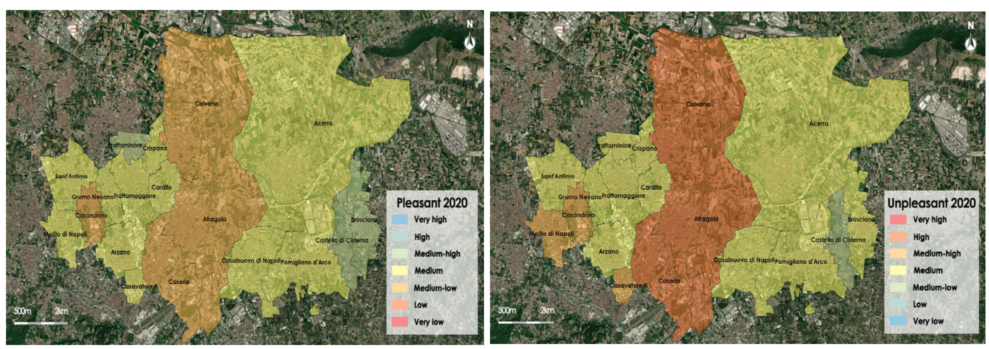
	Basic Emotions	Secondary Emotions
Pleasant	Expectation	Awe
	Joy	Love
	Surprise	Optimism
	Trust	Content
Unpleasant	Anger	Aggression
	Sadness	Disapproval
	Disgust	Remorse
	Fear	Submission

By assigning to each emotional category, all the phrases that were semantically related to it but had been reduced to their stemmed form, the dictionary of emotional words was generated in accordance with [12].

Following the Corpus Documents' processing, as outlined in the Term Filtering phase, terms were associated with the dictionary category when they were discovered there.

For each document, fuzzified pleasant and unpleasant relevance indices where a document corresponds to a specific province and time frame were calculated.

Figure 6 shows the thematic maps of the pleasant (Figure 6a) and unpleasant (Figure 6b) emotion categories' relevance measured in 2020.



(**a**) Pleasant emotion relevance (**b**) Unpleasant emotion relevance

Figure 6. Pleasant and unpleasant emotion relevance thematic maps in the year 2020.

The pleasant emotion relevance thematic map shows that in 2020 the municipalities of Brusciano, Castello di Cisterna, and Frattaminore were those in which pleasant emotional feelings prevailed, with the relevance of the pleasant emotion being Medium-high.

The unpleasant emotion relevance thematic map highlights that in 2020 the municipalities in which unpleasant emotions prevail are Caivano, Afragola, and Casoria, with High relevance, and Casandrino, Melito di Napoli and Casavatore, with Medium high relevance.

Table 2 shows in detail the membership degrees of each municipality to all pleasant emotion relevance fuzzy sets of the ECR fuzzy partition measured in 2020. The last column shows the relevance assigned to each municipality, given by the label of the fuzzy set with the greatest degree of membership.

In the municipalities of Brusciano, Castello di Cisterna, and Frattaminore, the pleasant relevance class with the highest membership degree (0.7, 0.7, and 0.5, respectively, highlighted in black) is Medium high. All other municipalities belong to the pleasant relevance classes Medium or Medium low.

Table 2. Relevance of pleasant emotional categories in the year 2020.

Municipality	Very Low	Low	Medium Low	Medium	Medium High	High	Very High	Relevance
Acerra	0.00	0.00	0.00	0.90	0.10	0.00	0.00	Medium
Afragola	0.00	0.40	0.60	0.00	0.00	0.00	0.00	Medium low
Arzano	0.00	0.00	0.20	0.80	0.00	0.00	0.00	Medium
Brusciano	0.00	0.00	0.00	0.30	**0.70**	0.00	0.00	**Medium high**
Caivano	0.00	0.30	0.70	0.00	0.00	0.00	0.00	Medium low
Cardito	0.00	0.00	0.10	0.90	0.00	0.00	0.00	Medium
Casalnuovo di Napoli	0.00	0.00	0.00	0.70	0.30	0.00	0.00	Medium
Casandrino	0.00	0.00	0.60	0.40	0.00	0.00	0.00	Medium low
Casavatore	0.00	0.00	0.30	0.70	0.00	0.00	0.00	Medium
Casoria	0.00	0.00	0.95	0.05	0.00	0.00	0.00	Medium low
Castello di Cisterna	0.00	0.00	0.00	0.00	**0.70**	0.30	0.00	**Medium high**
Crispano	0.00	0.00	0.00	0.60	0.40	0.00	0.00	Medium
Frattamaggiore	0.00	0.00	0.15	0.85	0.00	0.00	0.00	Medium
Frattaminore	0.00	0.00	0.00	0.50	**0.50**	0.00	0.00	**Medium high**
Grumo Nevano	0.00	0.15	0.15	0.85	0.00	0.00	0.00	Medium
Melito di Napoli	0.00	0.30	0.30	0.70	0.00	0.00	0.00	Medium
Pomigliano d’Arco	0.00	0.20	0.20	0.80	0.00	0.00	0.00	Medium
Sant’Antimo	0.00	0.00	0.00	0.80	0.20	0.00	0.00	Medium

The detailed results for the unpleasant emotion relevance are shown in Table 3, where the membership degrees to each fuzzy set are shown for each municipality.

The municipalities of Afragola, Caivano, and Casoria belong to the unpleasant relevance class High, with membership degrees of 1.0, 1.0, and 0.9, respectively. The municipalities of Casandrino, Casavatore, and Melito di Napoli belong to the unpleasant relevance class Medium high with membership degrees of 0.8, 0.6, and 0.6, respectively. All other municipalities belong to the unpleasant relevance classes Medium or Medium low.

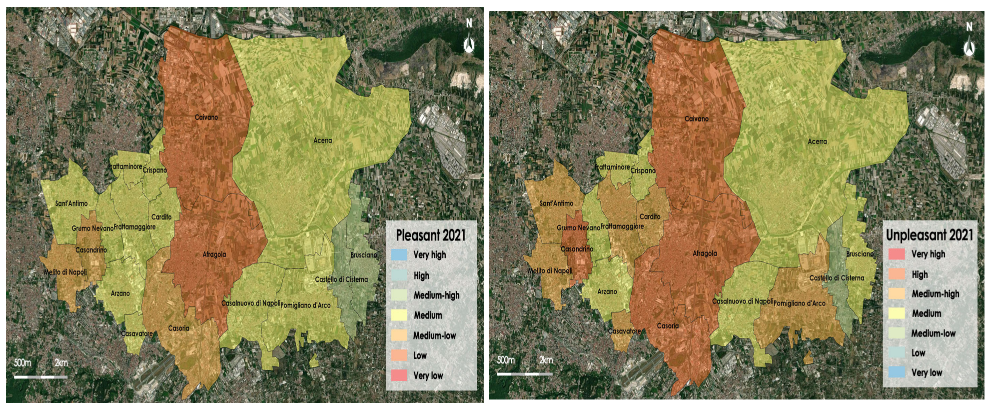
Table 3. Relevance of unpleasant emotional categories in the year 2020.

Municipality	Very Low	Low	Medium Low	Medium	Medium High	High	Very High	Relevance
Acerra	0.00	0.00	0.05	0.95	0.00	0.00	0.00	Medium
Afragola	0.00	0.00	0.00	0.00	0.00	**1.00**	0.40	**High**
Arzano	0.00	0.00	0.00	0.60	0.40	0.00	0.00	Medium
Brusciano	0.00	0.00	0.35	0.65	0.00	0.00	0.00	Medium
Caivano	0.00	0.00	0.00	0.00	0.00	**1.00**	0.30	**High**
Cardito	0.00	0.00	0.00	0.80	0.20	0.00	0.00	Medium
Casalnuovo di Napoli	0.00	0.00	0.15	0.85	0.00	0.00	0.00	Medium
Casandrino	0.00	0.00	0.00	0.00	**0.80**	0.20	0.00	**Medium high**
Casavatore	0.00	0.00	0.00	0.40	**0.60**	0.00	0.00	**Medium high**
Casoria	0.00	0.00	0.00	0.00	0.10	**0.90**	0.00	**High**
Castello di Cisterna	0.00	0.00	0.65	0.35	0.00	0.00	0.00	Medium low
Crispano	0.00	0.00	0.20	0.80	0.00	0.00	0.00	Medium
Frattamaggiore	0.00	0.00	0.00	0.70	0.30	0.00	0.00	Medium
Frattaminore	0.00	0.00	0.25	0.75	0.00	0.00	0.00	Medium
Grumo Nevano	0.00	0.00	0.00	0.70	0.30	0.00	0.00	Medium
Melito di Napoli	0.00	0.00	0.00	0.40	**0.60**	0.00	0.00	**Medium high**
Pomigliano d’Arco	0.00	0.00	0.00	0.60	0.40	0.00	0.00	Medium
Sant’Antimo	0.00	0.00	0.10	0.90	0.00	0.00	0.00	Medium

The results in Tables 2 and 3, showing the pleasant and unpleasant emotion relevance assigned in 2020 to each municipality, highlight:

- The presence of an urban area of greater comfort felt by the citizen, which includes the neighboring municipalities of Brusciano and Castello di Cisterna:
- A larger urban area that includes municipalities in the central strip where a greater discomfort felt by the citizens prevails;
- Another urban area in the southwest of the study area covers the municipalities of Casandrino, Grumo Nevano, and Melito di Napoli.

Figure 7 shows the thematic maps of the pleasant (Figure 7a) and unpleasant (Figure 7b) emotion categories' relevance measured in 2021.



(**a**) Pleasant emotion relevance (**b**) Unpleasant emotion relevance

Figure 7. Pleasant and unpleasant emotion relevance thematic maps in the year 2021.

The pleasant emotion relevance thematic map shows that in 2021 the municipalities of Brusciano and Castello di Cisterna are those in which pleasant emotional feelings prevailed, with the relevance of the pleasant emotion being Medium-high.

The unpleasant emotion relevance thematic map shows that in 2021 the municipalities in which unpleasant emotions prevail are Caivano, Afragola, Casoria, and Casandrino, with High relevance, and Sant'Antimo, Melito di Napoli, Cardito, Frattamaggiore, Casavatore, and Pomigliano d'Arco with Medium high relevance.

The detailed results for the pleasant and unpleasant emotion relevance are shown, respectively, in Tables 4 and 5, where, for each municipality, are shown the membership degrees to each fuzzy set.

Table 4 shows that the municipalities Brusciano and Castello di Cisterna belong to the pleasant relevance class Medium high with membership degrees of 0.5 and 0.9, respectively. All other municipalities belong to the pleasant relevance classes Medium or Medium low.

Table 4. Relevance of pleasant emotional categories in the year 2021.

Municipality	Very Low	Low	Medium Low	Medium	Medium High	High	Very High	Relevance
Acerra	0.00	0.00	0.10	0.90	0.00	0.00	0.00	Medium
Afragola	0.00	0.70	0.30	0.00	0.00	0.00	0.00	Low
Arzano	0.00	0.00	0.20	0.80	0.00	0.00	0.00	Medium
Brusciano	0.00	0.00	0.00	0.50	**0.50**	0.00	0.00	**Medium high**
Caivano	0.00	0.50	0.50	0.00	0.00	0.00	0.00	Low
Cardito	0.00	0.00	0.25	0.75	0.00	0.00	0.00	Medium
Casalnuovo di Napoli	0.00	0.00	0.15	0.85	0.00	0.00	0.00	Medium
Casandrino	0.00	0.00	0.75	0.25	0.00	0.00	0.00	Medium low
Casavatore	0.00	0.00	0.45	0.55	0.00	0.00	0.00	Medium

Casoria	0.00	0.10	0.90	0.00	0.00	0.00	0.00	Medium low
Castello di Cisterna	0.00	0.00	0.00	0.00	**0.90**	0.10	0.00	**Medium high**
Crispano	0.00	0.00	0.00	1.00	0.00	0.00	0.00	Medium
Frattamaggiore	0.00	0.00	0.30	0.70	0.00	0.00	0.00	Medium
Frattaminore	0.00	0.00	0.00	0.70	0.30	0.00	0.00	Medium
Grumo Nevano	0.00	0.00	0.20	0.80	0.00	0.00	0.00	Medium
Melito di Napoli	0.00	0.00	0.60	0.40	0.00	0.00	0.00	Medium low
Pomigliano d'Arco	0.00	0.00	0.35	0.65	0.00	0.00	0.00	Medium
Sant'Antimo	0.00	0.00	0.25	0.75	0.00	0.00	0.00	Medium

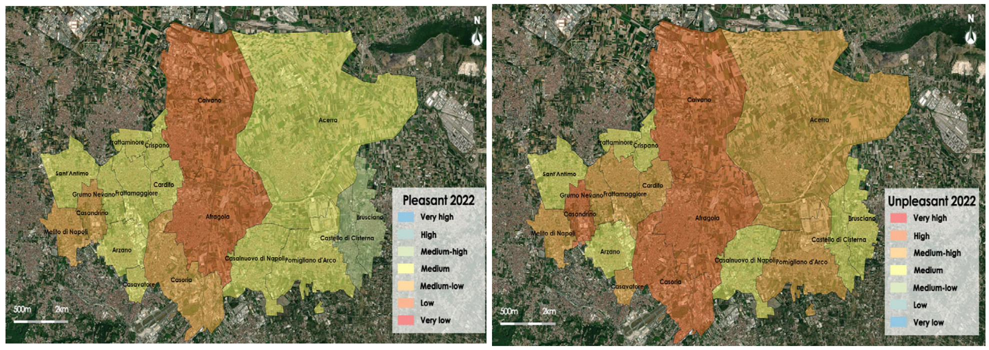
Table 5 shows that the municipalities Afragola, Caivano, Casandrino, and Casoria belong to the unpleasant class High with membership degrees of 1.0, 1.0, 0.5, and 1.0, respectively. The municipalities Cardito, Casavatore, Frattamaggiore, Melito di Napoli, Pomigliano d'Arco, and Sant'Antimo, belong to the unpleasant class Medium high with membership degrees of 0.5, 0.9, 0.6, 0.8, 0.7, and 0.5, respectively. All other municipalities belong to the pleasant relevance classes Medium or Medium low.

Table 5. Relevance of unpleasant emotional categories in the year 2021.

Municipality	Very Low	Low	Medium Low	Medium	Medium High	High	Very High	Relevance
Acerra	0.00	0.00	0.00	0.80	0.20	0.00	0.00	Medium
Afragola	0.00	0.00	0.00	0.00	0.00	**1.00**	0.70	**High**
Arzano	0.00	0.00	0.00	0.60	0.40	0.00	0.00	Medium
Brusciano	0.00	0.00	0.25	0.75	0.00	0.00	0.00	Medium
Caivano	0.00	0.00	0.00	0.00	0.00	**1.00**	0.50	**High**
Cardito	0.00	0.00	0.00	0.50	**0.50**	0.00	0.00	**Medium high**
Casalnuovo di Napoli	0.00	0.00	0.00	0.70	0.30	0.00	0.00	Medium
Casandrino	0.00	0.00	0.00	0.00	0.50	**0.50**	0.00	**High**
Casavatore	0.00	0.00	0.00	0.10	**0.90**	0.00	0.00	**Medium high**
Casoria	0.00	0.00	0.00	0.00	0.00	**1.00**	0.10	**High**
Castello di Cisterna	0.00	0.00	0.55	0.45	0.00	0.00	0.00	Medium low
Crispano	0.00	0.00	0.00	1.00	0.00	0.00	0.00	Medium
Frattamaggiore	0.00	0.00	0.00	0.40	**0.60**	0.00	0.00	**Medium high**
Frattaminore	0.00	0.00	0.15	0.85	0.00	0.00	0.00	Medium
Grumo Nevano	0.00	0.00	0.00	0.60	0.40	0.00	0.00	Medium
Melito di Napoli	0.00	0.00	0.00	0.00	**0.80**	0.20	0.00	**Medium high**
Pomigliano d'Arco	0.00	0.00	0.00	0.30	**0.70**	0.00	0.00	**Medium high**
Sant'Antimo	0.00	0.00	0.00	0.50	**0.50**	0.00	0.00	**Medium high**

Analyzing the results in Tables 4 and 5, showing the pleasant and unpleasant emotion relevance assigned in the year 2021 to each municipality, detected the presence of the areas of comfort and discomfort felt by citizens observed in 2020.

Figure 8 shows the thematic maps of the pleasant (Figure 8a) and unpleasant (Figure 8b) emotion categories' relevance measured in 2022.



(**a**) Pleasant emotion relevance (**b**) Unpleasant emotion relevance

Figure 8. Pleasant and unpleasant emotion relevance thematic maps in 2022.

The pleasant emotion relevance thematic map shows that in 2022 the municipalities of Brusciano and Castello di Cisterna are those in which pleasant emotional feelings prevailed, with the relevance of the pleasant emotion being Medium-high.

The unpleasant emotion relevance thematic map shows that in 2022 the municipalities in which unpleasant emotions prevail are Caivano, Afragola, Casoria, and Casandrino, with High relevance, and Acerra, Cardito, Casavatore, Frattamaggiore, Grumo Nevano, Melito di Napoli, and Pomigliano d'Arco with Medium high relevance.

The detailed results for the pleasant and unpleasant emotion relevance are shown, respectively, in Tables 6 and 7, where, for each municipality, the membership degrees to each fuzzy set are shown.

Table 6 shows that the municipalities Brusciano and Castello di Cisterna belong to the pleasant relevance class Medium high with membership degrees of 0.6 and 1.0, respectively. All other municipalities are assigned a pleasant relevance class Low, Medium low, or Medium.

Table 6. Relevance of pleasant emotional categories in the year 2022.

Municipality	Very Low	Low	Medium Low	Medium	Medium High	High	Very High	Relevance
Acerra	0.00	0.00	0.25	0.75	0.00	0.00	0.00	Medium
Afragola	0.00	0.80	0.20	0.00	0.00	0.00	0.00	Low
Arzano	0.00	0.00	0.20	0.80	0.00	0.00	0.00	Medium
Brusciano	0.00	0.00	0.00	0.40	**0.60**	0.00	0.00	**Medium high**
Caivano	0.00	0.70	0.30	0.00	0.00	0.00	0.00	Low
Cardito	0.00	0.00	0.40	0.60	0.00	0.00	0.00	Medium
Casalnuovo di Napoli	0.00	0.00	0.10	0.90	0.00	0.00	0.00	Medium
Casandrino	0.00	0.00	0.90	0.10	0.00	0.00	0.00	Medium low
Casavatore	0.00	0.00	0.40	0.60	0.00	0.00	0.00	Medium
Casoria	0.00	0.00	1.00	0.00	0.00	0.00	0.00	Medium low
Castello di Cisterna	0.00	0.00	0.00	0.00	**1.00**	0.00	0.00	**Medium high**
Crispano	0.00	0.00	0.15	0.85	0.00	0.00	0.00	Medium
Frattamaggiore	0.00	0.00	0.25	0.75	0.00	0.00	0.00	Medium
Frattaminore	0.00	0.00	0.00	0.80	0.20	0.00	0.00	Medium
Grumo Nevano	0.00	0.00	0.30	0.70	0.00	0.00	0.00	Medium
Melito di Napoli	0.00	0.00	0.55	0.45	0.00	0.00	0.00	Medium low
Pomigliano d'Arco	0.00	0.00	0.50	0.50	0.00	0.00	0.00	Medium
Sant'Antimo	0.00	0.00	0.20	0.80	0.00	0.00	0.00	Medium

Table 7 shows that the municipalities Afragola, Caivano, Casandrino, and Casoria belong to the unpleasant relevance class High with membership degrees of 1.0, 1.0, 0.8, and 1.0, respectively. The municipalities Melito di Napoli, Grumo Nevano, Cardito, Frattamaggiore, Casavatore, Acerra, and Pomigliano d'Arco belong to the unpleasant relevance class Medium high with membership degrees of 0.5, 0.8, 0.8, 0.5, 0.6, 0.9, and 1.0, respectively. The unpleasant relevance class Medium is assigned to all other municipalities.

Table 7. Relevance of unpleasant emotional categories in 2022.

Municipality	Very Low	Low	Medium Low	Medium	Medium High	High	Very High	Relevance
Acerra	0.00	0.00	0.00	0.50	**0.50**	0.00	0.00	**Medium high**
Afragola	0.00	0.00	0.00	0.00	0.00	**1.00**	0.80	**High**
Arzano	0.00	0.00	0.00	0.60	0.40	0.00	0.00	Medium
Brusciano	0.00	0.00	0.30	0.70	0.00	0.00	0.00	Medium
Caivano	0.00	0.00	0.00	0.00	0.00	**1.00**	0.70	**High**
Cardito	0.00	0.00	0.00	0.20	**0.80**	0.00	0.00	**Medium high**
Casalnuovo di Napoli	0.00	0.00	0.00	0.80	0.20	0.00	0.00	Medium
Casandrino	0.00	0.00	0.00	0.00	0.20	**0.80**	0.00	**High**
Casavatore	0.00	0.00	0.00	0.20	**0.80**	0.00	0.00	**Medium high**
Casoria	0.00	0.00	0.00	0.00	0.00	**1.00**	0.00	**High**
Castello di Cisterna	0.00	0.00	0.50	0.50	0.00	0.00	0.00	Medium
Crispano	0.00	0.00	0.00	0.70	0.30	0.00	0.00	Medium
Frattamaggiore	0.00	0.00	0.00	0.50	**0.50**	0.00	0.00	**Medium high**
Frattaminore	0.00	0.00	0.10	0.90	0.00	0.00	0.00	Medium
Grumo Nevano	0.00	0.00	0.00	0.40	**0.60**	0.00	0.00	**Medium high**
Melito di Napoli	0.00	0.00	0.00	0.00	**0.90**	0.10	0.00	**Medium high**
Pomigliano d'Arco	0.00	0.00	0.00	0.00	**1.00**	0.00	0.00	**Medium high**
Sant'Antimo	0.00	0.00	0.00	0.60	0.40	0.00	0.00	Medium

Also, in 2022, the two areas of high comfort and high discomfort felt by citizens are the areas including the municipalities of Brusciano and Castello di Cisterna, the central area, and the southwest area covering the municipalities of Casandrino, Grumo Nevano, and Melito di Napoli.

To detect hot and cold spots, we annotate as a cold or hot spot a municipality classified with pleasant or unpleasant relevance classes of Medium high, High, or Very high in all three consecutive years 2020, 2021, and 2022.

The thematic map in Figure 9 shows the hot and cold spots detected in the study area.

The northeastern area of the province of Naples is covered by two hot spots. The largest is located along the central area of the study area and covers the extension of the municipalities Afragola, Caivano, Casavatore, and Casoria; the other is located to the southwest and covers the extension of the municipalities Casandrino, Grumo Nevano, and Melito di Napoli. The study area is also covered by a cold spot located to the southwest and including the municipalities of Brusciano and Castello di Cisterna.

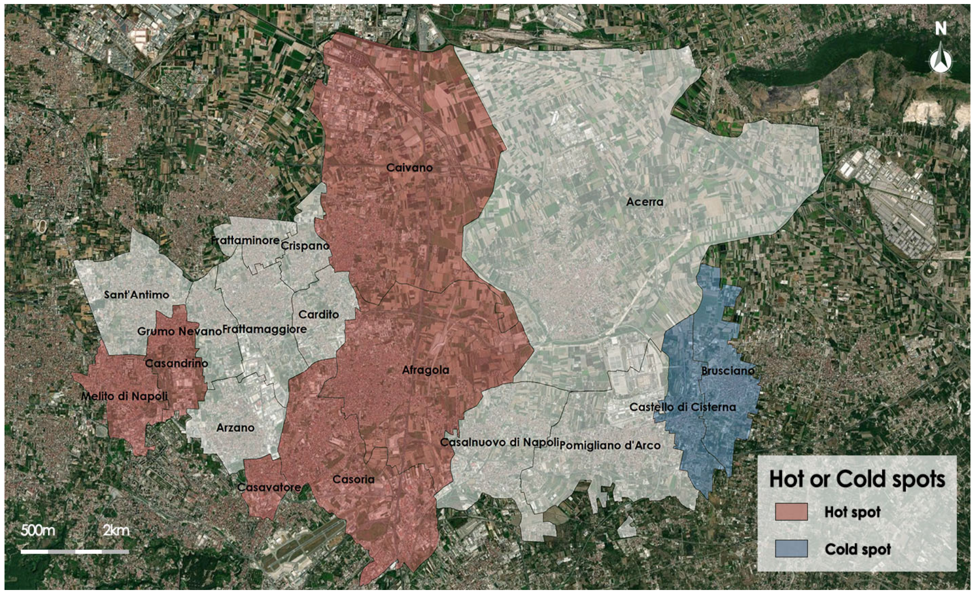


Figure 9. Map of Hot and Cold spots in the study area.

To verify whether the presence of hot/cold spots is related to a high/low housing/building density, the values of housing densities were calculated by the number of inhabitants per square kilometer in the areas covered by hot and cold spots.

Table 8 shows the extension in km^2, the population, and the mean population density in the hot and cold spots. The mean population density is given by the average of the population densities measured in the subzone covered by the hot/cold spot.

Table 8. Hot and cold spots detected.

Spot Type	Municipalities	Population	Area (km^2)	Mean Population Density
Hot spot	Afragola, Caivano, Casavatore, Casoria	189,037	58.80	5664.50
Hot spot	Casandrino, Grumo Nevano, Melito di Napoli	66,454	9.86	6541.00
Cold spot	Brusciano, Castello di Cisterna	23,443	9.54	2385.00

The data in Table 8 demonstrates that the mean population density in the two hot spots is higher than the mean population density in the research region. This is because the mean population density in all subzones of the study area is 5008.54 residents per km^2. In contrast, the cold spot's mean population density is lower than the study area's mean population density. This supports the hypothesis that the discomfort felt by the resident population in the summer months is felt more in areas with a high housing/building density. Conversely, this discomfort is less felt in areas with lower housing/building density.

These results highlight the usefulness of our approach for the decision maker in identifying the most critical areas and those less exposed to the phenomenon and in evaluating any correlations with the urban characteristics of these areas.

5. Conclusions

We present a new method implemented in a GIS-based framework to detect hot and cold spots from social streams. A lightweight document emotion classification method is applied to measure the relevance of pleasant and unpleasant emotions in documents where a document is related to a subzone and a time frame. A fuzzy partition is used to fuzzify pleasant and unpleasant emotions' relevance. Subzones classified with pleasant (unpleasant) emotion relevance over a prefixed threshold are labeled as cold (hot) spots in all time frames.

We have applied the FESC framework to an urban study area given by the municipalities of the northeastern area of the province of Naples (Italy), analyzing the citizens' discomfort due to the presence of heat wave phenomena in the summer seasons of the years from 2020 to 2022. The findings indicate that the study area's two hot spots are located in areas with large densities of people and buildings. On the other hand, the identified cold patch is marked by an urban region with a lower population density than the research area's average.

The results obtained confirm the effectiveness of our approach, which detects hot and cold spots on the basis of the hidden knowledge inserted in social networks connected to the perception of comfort and discomfort of citizens in the presence of the analyzed phenomenon. The hot and cold spot detection algorithms in the literature only use measurement or collected data, in some cases insufficient to represent the critical issues actually perceived and present in the area.

In the future, we plan to use our methodology in various settlements to analyze the discomfort of citizens in the presence of various types of climatic and environmental phenomena. In addition, we intend to compare our method with the traditional hot and cold spot detection algorithms by integrating datasets consisting of geo-referenced event points with hidden knowledge extracted from social networks.

Author Contributions: Conceptualization, B.C., F.D.M., and V.M.; methodology, B.C., F.D.M., and V.M.; software, B.C., F.D.M. and V.M.; validation, B.C., F.D.M. and V.M.; formal analysis, B.C., F.D.M., and V.M.; investigation, B.C., F.D.M., and V.M.; resources, B.C., F.D.M., and V.M.; data curation, B.C., F.D.M., and V.M.; writing—original draft preparation, B.C., F.D.M., and V.M.; writing—review and editing, B.C., F.D.M., and V.M.; visualization, B.C., F.D.M. and V.M.; supervision, B.C., F.D.M., and V.M. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: Data sharing is not applicable.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Peng, S.; Cao, L.; Zhou, Y.; Ouyang, Z.; Yang, A.; Li, X.; Jia; W.; Shui, Yu. A survey on deep learning for textual emotion analysis in social networks. *Digit. Commun. Netw.* **2021**, *8*, 745–762. https://doi.org/10.1016/j.dcan.2021.10.003.
2. Adadi, A. A survey on data-efficient algorithms in big data era. *J. Big Data* **2021**, *8*, 54. https://doi.org/10.1186/s40537-021-00419-9.
3. van Engelen, J.E.; Hoos, H.H. A survey on semi-supervised learning. *Mach. Learn.* **2020**, *109*, 373–440. https://doi.org/10.1007/s10994-019-05855-6.
4. Triguero, I.; García, S.; Herrera, F. Self-labeled techniques for semi-supervised learning: Taxonomy. software and empirical study. *Knowl. Inf. Syst.* **2013**, *42*, 245–284. https://doi.org/10.1007/s10115-013-0706-y.
5. Aggarwal, C.C. A Survey of Stream Clustering Algorithms. In *Data Clustering. Algorithms and Applications*, 1st ed.; Aggarwal, C.C., Reddy, C.K., Eds.; Chapman and Hall/CRC: London, UK, 2014; pp. 652. ISBN 9781466558212.
6. Di Martino, F.; Senatore, S.; Sessa, S. A lightweight clustering-based approach to discover different emotional shades from social message streams. *Int. J. Intell. Syst.* **2019**, *34*, 1505–1523. https://doi.org/10.1002/int.22105.
7. Bezdek, J.C. *Pattern Recognition with Fuzzy Objective Function Algorithms*; Plenum Press: New York, NY, USA, 1981; pp. 272. ISBN: 978-0306406713.
8. Bezek, J.C.; Ehrlich, R.; Full, W. The fuzzy C-means Clustering Algorithm. *Comput. Geosci.* **1984**, *10*, 191–203. https://doi.org/10.1016/0098-3004(84)90020-7.
9. Kaymak, U.; Setnes, M. Fuzzy clustering with volume prototype and adaptive cluster merging. *IEEE Trans. Fuzzy Syst.* **2002**, *10*, 705–712. https://doi.org/10.1109/TFUZZ.2002.805901.
10. Cardone, B.; Di Martino, F.; Sessa, S. GIS-based fuzzy sentiment analysis framework to classify urban elements according to the orientations of citizens and tourists expressed in social networks. *Evol. Intell.* **2022**, *15*, 1959–1968. https://doi.org/10.1007/s12065-021-00603-z.
11. Cardone, B.; Di Martino, F.; Senatore, S. Improving the emotion-based classification by exploiting the fuzzy entropy in FCM clustering. *Int. J. Intell. Syst.* **2021**, *36*, 6944–6967. https://doi.org/10.1002/int.22575.
12. Cardone, B.; Di Martino, F.; Senatore, S. A fuzzy partition-based method to classify social messages assessing their emotional relevance. *Inf. Sci.* **2022**, *594*, 60–75. https://doi.org/10.1016/j.ins.2022.02.028.

13. Cardone, B.; Di Martino, F. A GIS-Based Fuzzy Multiclassification Framework Applied for Spatiotemporal Analysis of Phenomena in Urban Contexts. *Information* **2022**, *13*, 248;. https://doi.org/10.3390/info13050248.
14. Getis, A.; Ord, J.K. The Analysis of Spatial Association by Use of Distance Statistics. *Geogr. Anal.* **1992**, *24*, 189–206. https://doi.org/10.1111/j.1538-4632.1992.tb00261.x.
15. Anselin, L. Local Indicators of Spatial Association-LISA. *Geogr. Anal.* **1995**, *27*, 93–115. https://doi.org/10.1111/j.1538-4632.1995.tb00338.x.
16. Samiul Islam, S.M.; Ashraful Islam, K.M.; Akter Mullick, M.R. Drought hot spot analysis using local indicators of spatial autocorrelation: An experience from Bangladesh. *Environ. Chall.* **2022**, *6*, 100410. https://doi.org/10.1016/j.envc.2021.100410.
17. Chainey, S.; Tompson, L.; Uhlig, S. The Utility of Hotspot Mapping for Predicting Spatial Patterns of Crime. *Secur. J.* **2008**, *21*, 4–28. https://doi.org/10.1057/palgrave.sj.8350066.
18. Devroye, L.; Rugosi, G. *Combinatorial Methods in Density Estimation*, 2001st ed.; Springer Series in Statistics; Springer: Berlin/Heidelberg, Germany, 2001; p. 208. ISBN 978-0387951171.
19. MacQueen, J.B. Some Methods for Classification and Analysis of Multivariate Observations. In *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability*; Le Cam, L.M., Neyman, J., Eds.; University of California Press: Oakland, CA, USA, 1967; Volume 1, pp. 281–297.
20. Agarval, J.; Nagpal, R.; Sehgal, R. Crime Analysis Using K-Means Clustering. *Int. J. Comput. Appl.* **2013**, *83*, 4.
21. Sing, A.K.; Manimannan, G. Detecting Hot Spots on Crime Data Using Data Mining and Geographical Information System. *Int, J. Stat. Math.* **2013**, *8*, 5–9.
22. Hajela, G.; Chawla, M.; Rasool, A. A Clustering Based Hot Spot Identification Approach for Crime Prediction. *Procedia Comput. Sci.* **2020**, *167*, 1462–1470.
23. Vadrevu, K.V.; Csiszar, I.; Ellicott, E.; Giglio, L.; Badarinath, K.V.S.; Vermote, E.; Justice, C. Hot Spot Analysis of Vegetation Fires and Intensity in the Indian Region. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2013**, *6*, 224–228.
24. Khairani, N.A.; Sutoyo, E. Application of K-Means Clustering Algorithm for Determination of Fire-Prone Areas Utilizing Hot Spots in West Kalimantan Province. *Int. J. Adv. Data Inf. Syst.* **2020**, *1*, 9–16.
25. Kaufman, L.; Rousseeuw, P.J. *Finding Groups in Data: An Introduction to Cluster Analysis*, 2nd ed.; John Wiley & Sons: Hoboken, NJ, USA, 2005; p. 342. ISBN 978-0471735786.
26. Hardika, E.; Atmaja, S. Implementation of k-Medoids Clustering Algorithm to Cluster Crime Patterns in Yogyakarta. *Int. J. Appl. Sci. Smart Technol.* **2019**, *1*, 38–48. https://doi.org/10.24071/ijasst.v1i1.1859.
27. Tabarej, M.S.; Minz, S. Rough-Set Based Hot Spot Detection in Spatial Data. In *Advances in Computing and Data Sciences*; ICACDS 2019: Communications in Computer and Information Science; Singh, M., Gupta, P., Tyagi, V., Flusser, J., Ören, T., Kashyap, R., Eds.; Springer: Singapore, 2019; Volume 1046, pp. 356–368. https://doi.org/10.1007/978-981-13-9942-8_34.
28. Havens, T.C.;Bezdek, J. C.; Leckie, C-; Hall, L. O.; Palaniswami, M. Fuzzy c-Means Algorithms for Very Large Data. *IEEE Trans. Fuzzy Syst.* **2012**, *20*, 1130–1146, doi: 10.1109/TFUZZ.2012.2201485.
29. Ansari, M.Y.; Prakash, A. Application of Spatio-Temporal Fuzzy C-Means Clustering for Crime Spot Detection. *Def. Sci. J.* **2018**, *68*, 374–380. https://doi.org/10.14429/dsj.68.12518.
30. Win, K.N.; Chen, J.; Chen, Y.; Fournier-Viger, P. PCPD: A Parallel Crime Pattern Discovery System for Large-Scale Spatiotemporal Data Based on Fuzzy Clustering. *Int. J. Fuzzy Syst.* **2019**, *21*, 1961–1974. https://doi.org/10.1007/s40815-019-00673-3.
31. Bandyopadhyaya, R.; Mitra, S. Fuzzy Cluster–Based Method of Hot Spot Detection with Limited Information. *J. Transp. Saf. Secur.* **2015**, *7*, 307–323. https://doi.org/10.1080/19439962.2014.959583.
32. Huang, Z.; Gao, S.; Cai, C.; Zheng, H. Pan, Z.; Li, W. A rapid density method for taxi passengers hot spot recognition and visualization based on DBSCAN. *Sci. Rep.* **2021**, *11*, 9420. https://doi.org/10.1038/s41598-021-88822-3.
33. Kumar, K.M.; Reddy, A.R.M. A fast DBSCAN clustering algorithm by accelerating neighbor searching using groups method. *Pattern Recognit.* **2016**, *58*, 39–48. https://doi.org/10.1016/j.patcog.2016.03.008.
34. Das, A.K.; Bhuyan, P.K. Self-Organizing Tree Algorithm (SOTA) Clustering for Defining Level of Service (LOS) Criteria of Urban Streets. *Period. Polytech. Transp. Eng.* **2019**, *47*, 309–317. https://doi.org/10.3311/PPtr.9911.
35. Cardone, B.; Di Martino, F. Fuzzy-Based Spatiotemporal Hot Spot Intensity and Propagation—An Application in Crime Analysis. *Electronics* **2022**, *11*, 370. https://doi.org/10.3390/electronics11030370.
36. Di Martino, F.; Sessa, S.; Barillari, E.S.; Barillari, M.S. Spatio-temporal hot spots and Application on a Disease Analysis Case via GIS. *Soft Comput.* **2014**, *18*, 2377–2384. https://doi.org/10.1007/s00500-013-1211-7.
37. Salton, G.; Buckley, C. Term-weighting approaches in automatic text retrieval. *Inf. Process. Manag.* **1988**, *24*, 513–523. https://doi.org/10.1016/0306-4573(88)90021-0.
38. Chakravorty, S. Identifying crime clusters: The spatial principles. *Middle States Geogr.* **1995**, *28*, 53–58.
39. Plutchik, R. *A General Psychoevolutionary Theory of Emotion*; Theories of Emotion; Plutchik, R., Kellerman, H. ,Eds.; Academic Press: Cambridge, MA, USA, 1980; pp. 3–33. https://doi.org/10.1016/B978-0-12-558701-3.50007-7.