

Data Mining Approach in Climate Classification and Climate Network Construction – Case Study Montenegro

Savo TOMOVIĆ, Predrag STANIŠIĆ, Srdjan KADIĆ

Abstract: In this study, we present results of applying data mining techniques on meteorological dataset obtained from the Institute of Hydrometeorology and Seismology of Montenegro. The dataset covers the measurements taken from all 11 main meteorological stations in Montenegro for the period 2010-2015. We build new climate classification system based on decision tree. The system is simpler (i.e. uses fewer attributes) and more accurate than the well-known Köppen climate classification system. In addition, we propose a novel procedure for climate network construction. Finally, we identify the regions within the same climate type in Montenegro's climate network with the Girvan-Newman algorithm for community detection and achieve better results with respect to classical K-means and hierarchical clustering algorithms.

Keywords: clustering; climate networks; Koppen climate classification system; mining meteorological data

1 INTRODUCTION

With the growing use of information technology, web systems and recent advances in database management systems used for creating, retrieving, updating and managing data, as well as fast and secure access to data repositories through computer networks and grids, the amount of data available to numerous companies, agencies and scientific laboratories has increased exponentially [1]. Naturally, all users want to utilize these huge repositories in order to improve their business activities.

As a result, data mining has become a very popular topic/technique/method for solving the "data rich and information poor" syndrome.

Data mining refers loosely to the process of semi-automatic, reliable and intelligent analysis of large sets of raw data and discovery of useful knowledge, information, instructions and answers [2]. Data mining techniques are getting increasingly integrated in day-to-day business operations making them more efficient and cheaper. At the same time, data mining is getting increasingly interesting for research workers and scientists from different areas. Consequently, real-world problems and new data mining topics are constantly growing.

In the last decades, extremely large amounts of climate data have become available in electronic form. It has become increasingly important for scientists, agriculturists, farmers, global food security, disaster management and related organizations to find an effective and accurate tool to analyze and extract hidden patterns from this huge pool/stock of data.

Actually, climate science is set to become one of the largest sources of data for data-driven research. The size of climate data in 2010 was estimated at 10 Petabytes (1 PB = 1,000 TB). This number is projected to grow exponentially to about 350 Petabytes by 2030 [1].

Many researchers have tried to use data mining techniques in this area. For example, Artificial Neural networks, k-Nearest Neighbor and Decision Trees were used for prediction of daily temperature and rainfall/precipitation prediction. Also, clustering techniques were used to categorize some areas into regions with similar climatic characteristics [3].

In this paper, the analyzed area is Montenegro. We analyze meteorological dataset obtained from the Institute of Hydrometeorology and Seismology of Montenegro. The dataset covers measurements for all 11 main meteorological stations in Montenegro for the period 2010-2015.

Based on this dataset, we developed a new climate classification system. Decision tree technique was applied. The system is simpler (i.e. uses fewer attributes) and it is more accurate than the well-known Köppen classification system. To the best of our knowledge, this kind of analysis has not been performed yet.

In addition, we constructed the network of Montenegro climate stations. The extension of complex network theory [2] to climate sciences is a very recent area that yielded *climate networks*. They usually rely on gridded time series of meteorological preprocessed variables. There are many works where the mentioned variables, such as surface air temperature [4, 5], equipotential heights [6, 7], sea surface temperature [8, 9] and related data are analyzed.

Climate network is usually constructed in the following procedure [10]. Single points - stations of measurement, identify the nodes of the network. Each node has a variable whose value varies over time. Once the time series of the variable is obtained for each node, the cross-correlation between all pairs of them is evaluated. A link between two nodes exists if there is a significant statistical interdependence between their time series. Typically, the linear cross-correlation function is used as the simplest measure of the statistical interdependence of temporal series [11].

The Montenegrin climate network contains 11 nodes. According to our procedure, a link between two nodes exists if they have obtained the same Köppen climate type at least once in the period 2010-2015. In that way, all available attributes are exploited in network construction process, not just one variable (for example air temperature or precipitation). At the same time, the selection of the threshold τ is avoided, since it is a non-trivial aspect of building a climate network (link between nodes exists if absolute value of correlation coefficient is greater than τ). Also, the selection of an association measure of the statistical interdependence of the temporal series (linear or nonlinear) is avoided.

2 DATASET DESCRIPTION AND PRE-PROCESSING

Montenegro (meaning "Black Mountain") is located in the southern part of the Balkan Peninsula. To the west, north and east it borders with Croatia (14 km), Bosnia-Herzegovina (225 km), Serbia and Kosovo (203 km), and to the south, it borders with Albania (172 km). Montenegro is a maritime country, with 293 km of the Adriatic coastline. Montenegro's territory measures 13,812 square km [12].

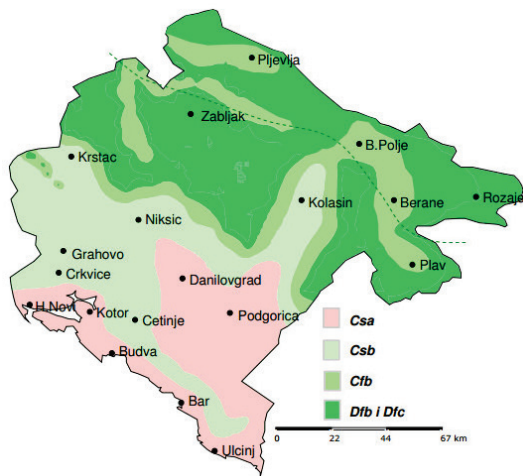


Figure 1 Climate types in Montenegro according to Köppen

Our study uses data obtained from the Institute of Hydrometeorology and Seismology of Montenegro for the period of 2010-2015.

Our dataset covers all 11 main meteorological stations in Montenegro: Bar, Berane, Bijelo Polje, Cetinje, Herceg Novi, Kolasin, Niksic, Podgorica, Pljevlja, Ulcinj, Zabljak. They are shown in Figure 1, along with the Köppen climate classification of Montenegro. More details about each climate type are given in Tab. 1 [13].

The dataset contains the following instrumental measurements: air temperature, extreme air temperatures, minimal temperature on 5 cm above the soil, soil temperatures on standard depths, air pressure, pressure tendency, air humidity, precipitation height and intensity, wind direction and speed and duration of Sun/solar radiation. The following visual observations are registered: cloudiness (quantity, kind and altitude of lower cloud basis), horizontal visibility, precipitation hydrometeor kind/type (rain, snow, hail, dew, hoarfrost etc.), soil state and phenological observations.

Köppen's classification is based on a subdivision of terrestrial climates into five major types. They are represented by capital letters A, B, C, D and E. Each of these climate types except for B is defined by temperature criteria. Type B designates climates in which the controlling factor on vegetation is dryness (rather than coldness) [14].

The Köppen climate classification uses monthly temperature and precipitation averages for the twelve months, usually averaged over a long period. Climate types are represented by a two or three letter combination in which the first letter defines the major type. The major types can be further divided into subtypes based on the

precipitation pattern (second letter, except for the E type) and the temperature (third letter) [15].

As noted above, temperature defines four major climate types. Additional letters designate the various subtypes. Type A climates (the warmest) are differentiated based on the seasonality of precipitation: Af (no dry season), Am (short dry season) or Aw (winter dry season). Type E climates (the coldest) are conventionally separated into tundra (ET) and snow/ice climates (EF). The mid-latitude C and D climates are given the second letter: f (no dry season), w (winter dry) or s (summer dry). The third symbol indicates the warmth of the summer or the coldness of the winter: a, b, c, or d. The last subclass exists only for D climates [14].

Table 1 Köppen climate type definition

Köppen climate type			Station
C (temperature of the warmest month $\geq 10^{\circ}\text{C}$, and temperature of the coldest month $< 18^{\circ}\text{C}$ but $> -3^{\circ}\text{C}$)	s (precipitation in the driest month of summer half of the year is less than 30 mm and less than one-third of the wettest month of the winter half)	a (temperature of the warmest month $\geq 22^{\circ}\text{C}$) b (temperature of each of four warmest months $\geq 10^{\circ}\text{C}$ but the warmest month $< 22^{\circ}\text{C}$)	Podgorica, H. Novi, Bar, Ulcinj
	f (precipitation more evenly distributed throughout year)	b (temperature of each of four warmest months $\geq 10^{\circ}\text{C}$ the warmest month $< 22^{\circ}\text{C}$)	Berane, Bijelo Polje, Pljevlja
	D (temperature of warmest month $\geq 10^{\circ}\text{C}$, and temperature of the coldest month $< -3^{\circ}\text{C}$)	f (precipitation more evenly distributed throughout year)	b and c

Bearing in mind that Köppen classification system uses just air temperature and precipitation averages, the original dataset is projected on these variables.

Missing values were estimated by using the attribute mean for all samples belonging to the same class as the given tuple [16]. For example, if air temperature for August 2010 in Bar is unknown, this value is estimated by average temperatures in August 2011-2015 in Bar.

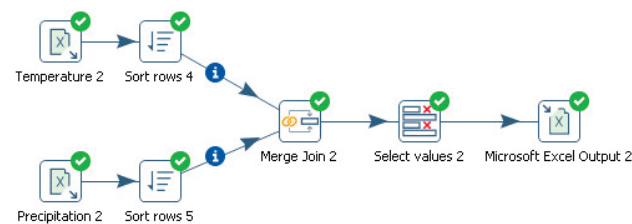


Figure 2 Pentaho transformation for selecting temperature and precipitation from the original dataset

All data pre-processing was done in Pentaho Data Integration software [17]. Fig. 2 depicts the transformation by which we obtained the dataset in Excel format with

temperature and precipitation values. We omit detailed explanations.

The resulting dataset contains 66 rows, 6 rows per each station. Each record contains the name of station, air temperature averaged from January to December and precipitation from January to December, in 25 fields altogether. Records are sorted and grouped by station names. The first record from the group refers to the year of 2010. The last record refers to the year of 2015. In the following sections, this dataset will be referred to as MNE climate dataset.

3 DATA MINING APPROACH IN CLIMATE CLASSIFICATION

Classification is the task of learning a target function f that maps each attribute set x to one of the predefined class labels y . The target function is also known as the classification model [18].

The classification model is useful for descriptive and predictive modeling.

In descriptive modeling, classification model may serve as an explanatory tool to distinguish between objects of different classes. Also, it can be useful for summarization of large datasets through describing them with several classification rules. For example, it would be useful for biologist to have a descriptive model that explains what features define a vertebrate as a mammal, reptile, bird, fish or amphibian.

Predictive models serve to predict the class label of unknown records. For example, classification model may be useful for biologist to determine the class to which the creature belongs.

In this section, we build new descriptive model for climate classification in Montenegro. The model is based on decision tree classifier, which is a widely used classification technique [18].

The decision tree has three types of nodes:

- 1) A root node, which has no incoming edges and more outgoing edges.
- 2) Internal nodes, each of which has exactly one incoming edge and at least two outgoing edges
- 3) Leaf nodes, each of which has exactly one incoming edge and no outgoing edges.

Each leaf node is assigned a class label. The non-leaf nodes contain attribute test condition to separate records with different characteristics (classes). For example, minimal temperature is used to separate C and D climate types in the Köppen classification model (Tab. 3).

Classifying a new record is straightforward once a decision tree is constructed. Starting from the root node, we apply the test condition to the record and follow the appropriate branch based on the outcome of the test. This will lead us either to another internal node or to a leaf node. In an internal node, a new test condition is applied. In a leaf node, the classification process terminates by assigning the leaf's class label to the record.

C4.5 builds decision tree from a set of training data using the concept of information entropy [18]. The training data is a set $S = \{s_1, s_2, \dots, s_n\}$ of already classified samples. Each sample s_i consists of a m - dimensional vector $(x_{i1}, x_{i2}, \dots, x_{im})$ where the x_{ij} represents attribute

values of the sample, as well as the class in which s_i falls [19].

At each node, C4.5 chooses the attribute that most effectively partitions the training records into purer subsets. The splitting criterion is the normalized information gain (difference in entropy). The attribute with the highest normalized information gain is chosen to make the decision. The C4.5 algorithm then recurs on the smaller subsets [19].

The following is a recursive definition of the algorithm [19].

Step 1. If all the samples belong to the same class, the algorithm creates a leaf node labeled with that class.

Step 2. An attribute test condition is selected to partition the records into smaller subsets if they do not belong to the same class. A child node is created for each outcome of the test condition and the records are distributed to the children based on the outcomes. The algorithm is then recursively applied to each child node.

In this study, we use implementation of the C4.5 algorithm in the Weka data mining tool [19]. We build classifier that is simpler than the Köppen system because it uses fewer attributes. According to the Occam's razor (principle of parsimony), the simpler model is preferred over more complex model when it achieves the same or better accuracy, which is the case with our model [18].

As we stated before, the Köppen climate classification uses monthly temperature and precipitation averages for twelve months, averaged over some period. In this study, the period covers the years 2010 – 2015. We used Pentaho for averaging temperature and precipitation from MNE climate dataset for the specified period.

The result is the dataset with 11 rows, one per each station. Transformation is depicted in Fig. 3. The main step is named *Group by* (station name). For each station it calculates average column values for 2010-2015 years. The columns correspond to monthly temperature and precipitation averages.

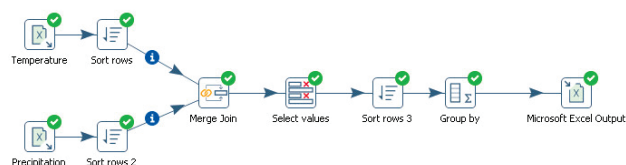


Figure 3 Averaging in Pentaho

Table 2 Variables used for Köppen climate type definition

Variable	Description
T_{min}	Temperature - minimum monthly value for the whole year
T_{max}	Temperature - maximum monthly value for the whole year
T_{mon}	Temperature - number of months satisfying criterion
P_{smin}	Precipitation - minimum monthly value for the summer months
P_{wmin}	Precipitation - minimum monthly value for the summer months
P_{wmax}	Precipitation - minimum monthly value for the summer months

The following schema used the same criteria for defining Köppen classifier as in [14, 15]. The main difference is that the schema is restricted to Montenegro area; for example, rules for major types A , B and E are ignored, as well as subtype w because they do not appear in Montenegro.

Table 3 Rules used for Köppen climate type definition

Type	Rule
First letter	C $-3\text{ }^{\circ}\text{C} < T_{\min} < 18\text{ }^{\circ}\text{C}$
	D $T_{\min} \leq -3\text{ }^{\circ}\text{C}$
Second letter	s $P_{\text{smin}} < P_{\text{wmin}}, P_{\text{wmax}} > 3P_{\text{smin}}$
	f Not f
Third letter	a $T_{\text{max}} \geq 22\text{ }^{\circ}\text{C}$
	b $T_{\text{max}} < 22\text{ }^{\circ}\text{C}, 4T_{\text{mon}} \geq 10\text{ }^{\circ}\text{C}$

Table 4 Köppen climate types for 2010-2015

Station	2010-2015
Bar	Csa
Berane	Cfb
Bijelo Polje	Csb
Cetinje	Csb
Herceg Novi	Csa
Kolasin	Csb
Nikšić	Csa
Podgorica	Csa
Pljevlja	Cfb
Ulcinj	Csa
Zabljak	Dsb

The criterion for each climate type is expressed using variables, which are defined in Tab. 2. The summer months on the Northern Hemisphere are April through September (AMJJAS). The winter months on the Northern Hemisphere are October through March (ONDJFM). For the Southern Hemisphere, it is vice versa [14, 15].

Tab. 3 contains rules for climate classification in Montenegro. In the last rule, $4T_{\text{mon}} \geq 10\text{ }^{\circ}\text{C}$ means "at least 4 months which mean temperature is above $10\text{ }^{\circ}\text{C}$ ".

Tab. 4 contains results of applying Köppen system to the averaged MNE climate dataset (Fig. 3). It can be concluded that Köppen type, with respect to [12, 13], has changed for the following stations: Bijelo Polje, Nikšić and Žabljak. It can be due to these regions being sensitive to climate variations, but this discussion is out of scope of this paper.

Now we will return to the original MNE climate dataset with 66 records. This dataset is used to assess performances of the Köppen classifier. Each record, representing the specific station in each respective year, is classified with the Köppen classification rules and the obtained class is compared with Tab. 4. For example, according to Tab. 4, Bar is labeled as Csa. At the same time, for Bar station the Köppen rules predict Csa climate type for all considered six years. Consider now Cetinje, classified as Csb in Tab. 4. The Köppen rules assign class Cfb to the years 2012 and 2015, while class Csb is assigned for the four other years.

Evaluation of the performance of a classification model is based on the counts of test records correctly predicted by the model [18]. Accuracy is performance metric, defined as follows:

$$\text{Accuracy} = \frac{\text{Number of correct predictions}}{\text{Total number of predictions}}$$

Most classification algorithms seek to develop models that attain highest accuracy [18].

The Köppen classification system achieves accuracy of 57.57% on the original MNE climate dataset. It makes 38 correct predictions along with the total of 66 predictions. Again, it can be due to significant climate

variations in the considered period, but this discussion is out of scope of this paper.

As we stated at the beginning of this section, we proceed towards providing an alternative climate classification system that uses fewer attributes and that is more accurate than the Köppen classification system. The system is the result of applying the C4.5 decision tree algorithm on the MNE climate dataset. For this purpose, we use WEKA package [19] and the J48 implementation of the algorithm C4.5.

The resulting tree is presented in Fig. 4. The tree defines climate classification system that uses four out of six variables from Tab. 2: T_{max} , T_{min} , T_{mon} and P_{wmax} . IF-THEN rules extracted from the tree are as follows:

```

IF (  $T_{\text{max}} \leq 23.2$ ,
    IF (  $T_{\text{mon}} \leq 4$ ,
        Dsb,
        IF (  $P_{\text{wmax}} \leq 202.8$ ,
            Cfb,
            IF (  $T_{\text{min}} \leq -2.1$ ,
                IF (  $T_{\text{mon}} \leq 5$ ,
                    Dsb,
                    Csb
                ),
                Csb
            )
        )
    )
    IF (  $T_{\text{min}} \leq 1$ ,
        IF (  $T_{\text{max}} \leq 24.3$ ,
            Csb,
            Csa
        ),
        Csa
    )
)
    
```

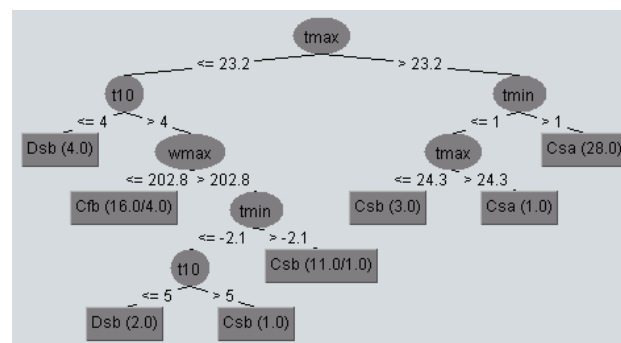


Figure 4 Decision tree for climate classification

Fig. 5 depicts the confusion matrix for our classification system [18]. Each entry f_{ij} in this matrix denotes the number of records from class i predicted to be of class j . Based on this, the total number of correct predictions made by the model is $\sum f_{ii}$. It gives accuracy of $47/66 = 71.21\%$, that is significantly higher than accuracy achieved with Köppen classification system.

a	b	c	d	<-- classified as
27	3	0	0	a = Csa
4	10	4	0	b = Csb
1	5	6	0	c = Cfb
0	2	0	4	d = Dsb

Figure 5 Confusion matrix

In order to show that the difference in accuracy in favor of our approach is statistically significant, we use McNemar’s test. This test has been used in different studies in the previous research in order to assess the performances of classification algorithms [20]. The test is a variant of χ^2 test and it is a non-parametric test. According to McNemar’s test we construct the following contingency table [21]:

Number of examples misclassified by both classifiers – n_{00}	Number of examples misclassified by A but not by B – n_{01}
Number of examples misclassified by B but not by A – n_{10}	Number of correctly classified examples by both classifiers – n_{11}

n_{00} denotes the number of instances when both algorithms failed and n_{11} denotes success for both algorithms. These two cases do not give much information about the algorithms’ performances as they do not indicate how their performances differ. However, the other two parameters (n_{01} and n_{10}) show cases where one of the algorithms failed and the other succeeded in indicating the performance discrepancies.

Under the null hypothesis, the two algorithms should have the same error rate, which means that $n_{01} = n_{10}$ [21]. The following statistics is approximately distributed as χ^2 with 1 degree of freedom [21]:

$$\frac{(|n_{01} - n_{10}| - 1)^2}{n_{01} + n_{10}}$$

If the null hypothesis is correct, then the probability that this quantity is greater than $\chi_{1, 0.95}^2 = 3.8414$ is less than 0.05. Contingency table in our case is as follows:

13	17
6	30

Inserting these values into the previous formula we obtain $4.348 > \chi_{1, 0.95}^2$. So, we may reject the null hypothesis in favor of the hypothesis that our classifier has better performance.

4 MONTENEGRO CLIMATE NETWORK

In this section, we propose a novel procedure for the climate network construction. In addition, we apply the Girvan-Newman algorithm for community detection on the previously built Montenegro climate network. Thus, we achieve better results in comparison to classical K-means clustering algorithm.

Complex network theory is a rapidly growing new field. Its fast development makes its application suitable to analyze a wide range of systems [2]. Here, we mention just a few of them: Internet and World Wide Web, neural connections, social media mining, human migration, cancer metastasis and earthquake occurrence [2]. Climate networks are an extension of the complex network theory to climate sciences.

Networks are usually modeled as undirected or directed graphs. The nodes represent entities, and an edge connects two nodes if the nodes are related by the relationship that characterizes the network.

Generally, nodes in the climate network are geographical locations where any pair of nodes is connected according to some rule. Actually, the edge existence between two nodes represents a statistical similarity between time-series that characterize that pair (e.g. linear correlation, mutual information, synchronization [10], etc.). In literature, there are many works where different meteorological variables, such as surface air temperature, equipotential heights, sea surface temperature, humidity and related data are analyzed.

In this paper we propose a novel procedure for network definition. Nodes are the main meteorological stations in Montenegro, so the Montenegro climate network has 11 nodes. For each node, monthly values from 2010 to 2015 are available for the following measurements: air temperature, extreme air temperatures, minimal temperature on 5 cm above the soil, soil temperatures on standard depths, air pressure, pressure tendency, air humidity, precipitation height and intensity, wind direction and speed and duration of Sun/solar radiation.

Instead of estimating the correlation coefficient between the time series of all pairs of nodes in order to define links, we classify each station according to the Köppen classification system. A station pair is considered connected if both stations get the same climate type in at least one year from the period 2010-2015. More formally, the link between stations P and Q exists if the following sentence is true:

$$\exists x, x \in [2010, 2015] \wedge climateType(P, x) = climateType(Q, x).$$

Again, we use the Pentaho Data Integration software to generate the pairs of connected stations. The transformation of the original MNE climate dataset is presented in Fig. 6. *Merge Join* step calculates *MNE climate dataset INNER JOIN MNE climate dataset* under the condition *ClimateType = ClimateType_1 and Year=Year_1*. As a result, we obtain CSV file with network graph representation where each line is unordered (undirected) pair (*StationP; StationQ*). The pair indicates link existence between P and Q . Fig. 7 depicts the Montenegro climate network graph after loading in Gephi [20].

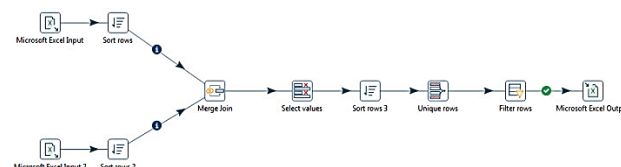


Figure 6 Pentaho transformation for generation of network graph

Next, we shall consider the clustering of the Montenegro climate graph as a way to identify communities. For this purpose, we use the Girvan-Newman algorithm [21]. It is based on finding the edges that are least likely to be inside a community.

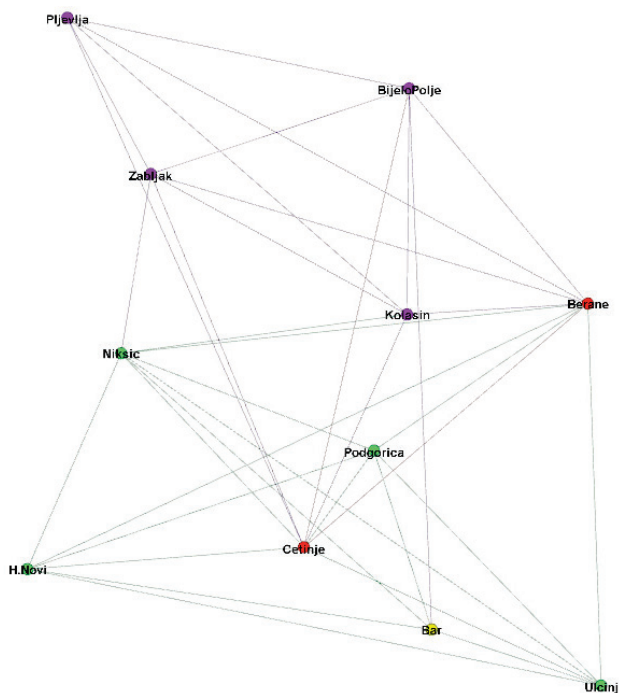


Figure 7 Montenegro climate network in Gephi

Define the *betweenness* of an edge (a, b) to be the number of pairs of nodes x and y such that the edge (a, b) lies on the shortest path between x and y [22]. Since there can be several shortest paths between x and y , the edge (a, b) is credited with the fraction of those shortest paths that include the edge (a, b) . High betweenness indicates that the edge (a, b) connects two different communities; that is, a and b do not belong to the same community.

The main idea of the Girvan Newman algorithm is expressed as a process of edge removal [22]. The algorithm starts with the graph and all its edges. It removes edges with the highest betweenness, until the graph breaks into a suitable number of connected components.

Fig. 7 also presents the community detection in the Montenegro climate network with the Girvan Newman algorithm. The algorithm recognized four communities indicated with different colors:

Yellow = {Bar}
 Green = {H.Novi, Niksic, Podgorica, Ulcinj}
 Red = {Berane, Cetinje}
 Purple = {Bijelo Polje, Kolasin, Pljevlja, Zabljak}.

To estimate the algorithm performances, we use *purity* [2]. For purity, we assume that the majority within the community determines the community climate type. For instance, stations in the Purple community have the following climate types respectively (Tab. 1): *Csb*, *Csb*, *Cfb*, *Dsb*. The majority consists of *Csb*, *Csb*, which means that Purple community is labeled with *Csb*.

Hence, we use the label of the community obtained in the above explained manner against the label of each member of that community to evaluate the algorithm. Formally,

$$Purity = \frac{1}{N} \sum_{i=1}^k \max_j |C_i \cap L_j|,$$

where k is the number of communities, N is the number of nodes in the network, L_j is the set of instances with label j in all communities, and C_i is the set of members in community i . In the case of the Montenegro climate network, purity is $\frac{1+4+1+2}{11} = 72.72\%$.

As a comparison, the K-Means clustering algorithm from the WEKA package correctly classified $\frac{1+4+1+1}{11} = 63.63\%$ stations in the Montenegro climate network, as Fig. 8 indicates. The hierarchical methods were even worse.

Classes to Clusters:

```
0 1 2 3 <-- assigned to cluster
0 4 1 0 | Csa
1 0 1 1 | Csb
1 0 1 0 | Cfb
1 0 0 0 | Dsb
```

```
Cluster 0 <-- Dsb
Cluster 1 <-- Csa
Cluster 2 <-- Cfb
Cluster 3 <-- Csb
```

Figure 8 K Means evaluation

5 CONCLUSION

In this study, we present the results of applying data mining techniques on meteorological dataset obtained from the Institute of Hydrometeorology and Seismology of Montenegro.

Montenegro climate dataset covers the measurements for all 11 main meteorological stations in Montenegro for the period [2010-2015].

We have built a new climate classification system based on decision tree algorithm. The system is simpler (i.e. uses fewer attributes) and it is more accurate than the well-known Köppen classification system.

In addition, we build the Montenegro climate network according to a novel procedure for climate network construction. Nodes are main meteorological stations in Montenegro. A station pair is considered as connected if both stations get the same climate type in at least one year from the period 2010-2015.

In order to confirm procedure applicability, we apply the Girvan-Newman algorithm for community detection on the Montenegro climate network and achieve better results in comparison to classical K-means and hierarchical clustering algorithms.

As for future, we plan to extend this study beyond analyzing just 11 main stations as Montenegro has a large number of meteorological stations that operate in accordance with the standards set by the World Meteorological Organization – WMO [12, 13]. In addition, we intend to significantly extend the considered period.

We hope that the previously described research will produce good results, giving us an opportunity to use the same techniques and successfully analyze some of the most important climate datasets mentioned in [15].

In addition, we plan to extend our research through coupling data mining techniques with geographic information system. The aim could be the construction of linked maps of geographic space.

6 REFERENCES

- [1] Faghmous, J. H. & Kumar, V. (2014). Spatio-temporal Data Mining for Climate Data: Advances, Challenges, and Opportunities. In: Chu, W. (eds) *Data Mining and Knowledge Discovery for Big Data. Studies in Big Data, vol. 1*. Springer, Berlin, Heidelberg, 83-116. https://doi.org/10.1007/978-3-642-40837-3_3
- [2] Zarafani, R.; Abbasi, M. A.; & Liu, H. (2014). *Social Media Mining*. Cambridge University Press, New York.
- [3] Scarsoglio, S., Laio, F., & Ridolfi, L. (2013). Climate Dynamics: A Network-Based Approach for the Analysis of Global Precipitation. *PLOS One*, 8(8), e71129. <https://doi.org/10.1371/journal.pone.0071129>
- [4] Wiedermann, M., Radebach, A., Donges, J. F., Kurths, J., & Donner, R. V. (2016). A climate network-based index to discriminate different types of El Niño and La Niña. *Geophysical Research Letters*, 43(16), 7176-7185. <https://doi.org/10.1002/2016GL069119>
- [5] Dodds, P. S., Mitchell, L., Reagan, A. J., & Danforth, C. M. (2016). Tracking Climate Change through the Spatiotemporal Dynamics of the Teletherms, the Statistically Hottest and Coldest Days of the Year. *PLOS One*, 11(5), e0154184. <https://doi.org/10.1371/journal.pone.0154184>
- [6] Tsonis, A. A., Wang, G., Swanson, K. L., Rodrigues, F. A., & Costa, L. F. (2011). Community structure and dynamics in climate networks. *Climate Dynamics*, 37(5), 933-940. <https://doi.org/10.1007/s00382-010-0874-3>
- [7] Donges, J. F., Schultz, H. C. H., Marwan, N., Zou, Y., & Kurths, J. (2011). Investigating the topology of interacting networks. *Eur. Phys. J.*, 84(4), 635-651. <https://doi.org/10.1140/epjb/e2011-10795-8>
- [8] Jung, W., Lee, D., & Kim, K. (2012). Characteristics of networks in sea surface temperatures. <http://adsabs.harvard.edu/abs/2012eguga..14.6771j>
- [9] Carpi, L., Masoller, C., Díaz-Guilera, A., & Ravetti, M. G. (2015). Dynamical analysis of the Indian Ocean climate network and its correlation with Australian Millennium Drought. <http://adsabs.harvard.edu/abs/2015EGUGA..17.7591C>
- [10] Steinhäuser, K., Chawla, N. V., & Ganguly, A. R. (2010). An exploration of climate data using complex networks. *ACM SIGKDD Explorations Newsletter*, 12(1), 25-32. <https://doi.org/10.1145/1882471.1882476>
- [11] Chen, Z., Hendrix, W., & Samatova, N. F. (2012). Community-based anomaly detection in evolutionary networks. *J. Intell. Inf. Syst.*, 39(1), 59-85. <https://doi.org/10.1007/s10844-011-0183-2>
- [12] Buric, D., Ducić, V., & Mihajlović, J. (2013). The Climate of Montenegro: Modifiers and Types – Part one. *Bulletin of Serbian Geographical Society*, 4, 84-102. <https://doi.org/10.2298/GSGD1304083B>
- [13] Buric, D., Ducić, V., & Mihajlović, J. (2014). The Climate of Montenegro: Modifiers and Types – Part two. *Bulletin of Serbian Geographical Society*, 1, 73-90. <https://doi.org/10.2298/GSGD1401073B>
- [14] Kottke, M., Grieser, J., Beck, C., Rudolf, B., & Rubel, F. (2006). World Map of the Köppen-Geiger climate classification updated. *Meteorologische Zeitschrift*, 15(3), 259-263. <https://doi.org/10.1127/0941-2948/2006/0130>
- [15] Chen, H. & Chen, D. (2014). *Köppen climate classification*. <http://hanschen.org/koppen>. (Accessed on 8.8.2014)
- [16] Han, J., Kamber, M., Kaufmann, M. (2011). *Data Preprocessing. Data Mining concepts and techniques*. 3rd Edition, MA/USA, 83-124.
- [17] <http://www.pentaho.com/product/data-integration>
- [18] Tan, P. N., Steinbach, M., Karpatne, A.; & Kumar, V. (2005). *Classification: Basic Concepts, Decision Trees, and Model Evaluation. Introduction to Data Mining*, 2nd Edition, 145-205.
- [19] <http://www.cs.waikato.ac.nz/ml/weka>
- [20] Dietterich, G. T. (1998). Approximate statistical tests for comparing supervised classification learning algorithms. *Neural Computation*, 10(7), 1895-1923. <https://doi.org/10.1162/089976698300017197>
- [21] Bostanci, B. & Bostanci, E. (2012). An Evaluation of Classification Algorithms Using Mc Nernan's Test. *Proceedings of Seventh International Conference on Bio-Inspired Computing: Theories and Applications (BIC-TA 2012)*, 201, 15-26.
- [22] <http://www.gephi.org>
- [23] Girvan, M. & Newman, M. E. J. (2002). Community structure in social and biological networks. *Proc. Natl. Acad. Sci. USA* 99, 7821-7826. <https://doi.org/10.1073/pnas.122653799>
- [24] Rajaraman, A. & Ullman, J. (2012). *Mining Social-Network Graphs. Mining of Massive Datasets*, 2nd edition. Cambridge University Press, New York, 343-405.

Contact information:

Savo TOMOVIĆ, PhD
University of Montenegro,
Džordža Vašingtona bb,
81000 Podgorica, Montenegro
savot@ac.me

Predrag STANIŠIĆ, PhD
University of Montenegro,
Džordža Vašingtona bb,
81000 Podgorica, Montenegro
pedjas@ac.me

Srdjan KADIĆ, PhD
University of Montenegro,
Džordža Vašingtona bb,
81000 Podgorica, Montenegro
skadic@ac.me