Prediction models and testing of resilience in regions: Covid19 economic impact in USA counties study case

Manousiadis, Charalampos; Gaki, Eleni

REGION
The Journal of ERSA
Powered by WU

Volume 10, Number 1, 2023, 113–132
DOI: 10.18335/region.v10i1.449

journal homepage: region.ersa.org
ISSN: 2409-5370

Prediction Models and Testing of Resilience in Regions: Covid19 Economic Impact in USA Counties Case Study

Charalampos Manousiadis[1], **Eleni Gaki**[1]

[1] University of the Aegean, Chios, Greece

Received: 19 July 2022/Accepted: 16 February 2023

Abstract. A significant amount of research has been conducted regarding the resilience of the regions and the factors that contribute to allow them to face challenges, crises, or disasters. The rise of promising sectors like Machine learning (ML) and Artificial Intelligence (AI) can enhance this research using computing power in regional economic, social, and environmental data analysis to find patterns and create prediction models. Through Machine Learning, the following research introduces the use of models that can predict the performance of a region in disasters. A case study of the performance of USA Counties during the Covid19 first wave period of the pandemic and the related restrictions that were applied by the authorities was used in order to reveal the obvious or hidden parameters and factors that affected their resilience, in particular their economic response, and other interesting patterns between all the involved attributes. This paper aims to contribute to a methodology and to offer useful guidelines in how regional factors can be translated and processed by data and ML/AI tools and techniques. The proposed models were evaluated on their ability to predict the economic performance of each county and in particular the difference of its unemployment rate between March and June of 2020. The former is based on several economic, social, and environmental data -up to that point in time- using classifiers like neural networks and decision trees. A comparison of the different models' execution was performed, and the best models were further analyzed and presented. Further execution results that identified patterns and connections between regional data and attributes are also presented. The main results of this research are i) a methodological framework of how regional status can be translated into digital models and ii) related examples of predictive models in a real case. An effort was also made to decode the results in terms of regional science to produce useful and meaningful conclusions, thus a decision tree is also presented to demonstrate how these models can be interpreted. Finally, the connection between this work and the strong current trend of regional and urban digitalization towards sustainability is established.

Key words: Regions, Resilience, Covid19, Machine Learning, Prediction Models, USA, Counties, Restrictions, Economic Impact, Unemployment rate

1 Introduction

Extensive research has been conducted regarding the resilience of the regions and the factors that contribute to allow them to face challenges, respond and recover from disturbances, crises, or disasters. There is a vast amount of related literature available.

This study aims to:

- create models through Machine Learning that predict the performance of a region in disasters to find the parameters/factors (confirm obvious from classic regional approaches or show hidden from data) that affect the resilience of the regions.
- mainly contribute to a methodology and offer useful guidelines on how regional factors can be translated and be processed by data and ML/AI tools and techniques.

1.1 The notion of resilience

The notion of resilience has recently gained more popularity due to the extended economic crises that most of the entire world faced in the last decades and the recent health crisis due to Covid19. Originally, resilience was used in engineering and ecology (Holling 1973), but since then, the concept has been used in many fields including regional economics. Half of the world's population resides in cities, with urban population expected to reach 70% of the global population by 2050 (United Nations 2016). Urban areas serve as locations that drive sustainable development, equality, inclusivity, cultural diversity, and are centers for innovation (Dhar, Khirfan 2017, Pickett et al. 2004). The recent economic and health crisis along with ecosystem pressure, climate change, migration, and other issues have increased the impact of urban crises. Therefore, the community resilience concept and corresponding mechanisms to build resilience on the community's complex systems have become popular (Abdul-Rahman et al. 2021). In addition, developing Community Resilience Assessment (CRA) tools is attributed to building a sustainable world (Seeliger, Turok 2013).

Regional resilience has been defined in many forms. A definition of regional resilience is the one proposed by Foster (2007). According to Foster (2007), it is the "ability of a region to anticipate, prepare for, respond to, and recover from disturbance". Bristow (2010)Bristow (2010) defines resilience as "the capacity of a system to absorb disturbance and reorganize while undergoing change, so as to still retain essentially the same function, structure and feedbacks". Kallioras (2011) argues that "resilience of a region is measured based on the evaluation of its ability to maintain a successful path of development after a disturbance, whether success is perceived in terms of traditional indicators such as growth or change of employment, or in terms of a synthetic index". According to Proag (2014), the concept of the regional resilience takes two broad forms: (1) hard resilience: the direct strength of structures when placed under pressure such as increasing the resilience of a structure through specific strengthening measures to reduce their probability of collapse, and (2) soft resilience: the ability of systems to absorb and recover from the impact of disruptive events without fundamental changes in function or structure, which depend on the flexibility and adaptive capacity of the system as a whole, rather than simply strengthening structures or institutions in relation to specific stresses, as in the hard resilience approach The most basic ways regions respond after each disorder are resistance, recovery, re-orientation, and renewal or resumption (Martin 2012). However, Pendall et al. (2010) argue that "regions face two main categories of disturbance: shocks and slow burns". In addition, according to the degree of resilience in a disturbance, regions are classified by three main categories (Briguglio et al. 2006, Hill et al. 2008):

- Economically resilient regions that improve or at least return to their original condition
- Shock-resistant regions that withstand and don't "escape" from their course
- Non-resilient regions which cannot return to their original state

The measurement of resilience is not an easy exercise, as this depends on the specific system under study and the ways that resilience is considered or requested to be calculated, either qualitatively or quantitatively. A qualitative assessment is useful to understand the current situation while quantitative measures give quantified estimates of performance that may be more meaningful to stakeholders e.g., policy makers seeking for parameters and values or researchers studying specific fields in the region (Proag 2014).

There are several different methodologies or complex indicators proposed in the literature or by authorities that measure resilience. They mainly involve economic, environmental, societal indicators, statistical analysis, and comparison with parameters strongly related to resilience, such as GDP and employment. The aim is to identify the drivers of crisis recovery and investigate the structural characteristics of the regions. Efforts are also made for a common framework. For example, a relevant technical report by the Joint Research Centre (JRC), which is the European Commission's science and knowledge service, proposes a simple 'handy' composite Regional Resilience Indicator to measure and monitor economic system resilience at the regional level in order to facilitate a common and easy understanding of this complex and dynamic process. This approach extends the existing theoretical framework and contributes to resilience a well-defined life cycle. The composite indicator weights have been attributed through weight elicitation techniques built upon principal component analysis (Serpieri, Pontarollo 2018).

In this study, the authors are exploring a hybrid (both qualitative and quantitative) assessment of the resilience. We are attempting to assess regional response to shock, classifying regions (counties) mainly in the range of the last two categories (shock resistant and non-resilient) without absolute correspondence, while also exploring the parameters that may affect this assessment and classification, which can then be used for policy making or research.

1.2 Machine Learning and Resilience

The rise of promising sectors in computer science such as Machine learning (ML) and Artificial Intelligence (AI) can boost many fields of research from e.g., medical applications and diagnosis (Shehab et al. 2022, Ahsan, Siddique 2022, Qezelbash-Chamak et al. 2022), to drug discovery (Patel, Shah 2022), and cybersecurity (Berghout et al. 2022). This also includes topics in the general framework of regional science, such as construction and infrastructure applications or seismic performance (Mirzaei et al. 2022, Mangalathu et al. 2022), regional crop yield forecasting (Paudel et al. 2022), spatio-temporal modeling of urban growth (Kim et al. 2022), and visual analyses of regional economy (Bai et al. 2022). The use of ML and the increasing computing power can support regional research to expand beyond the classic math, quantitative methods, and statistical analysis. It can contribute to the automation of searching, creating, calculating, and validating models, though hidden paths and by performing correlations and combinations which their execution would consume unrealistic time with the classic manual tools. Thus, ML can be applied in regional -economic, social, and environmental- data, to find patterns, forecast, and develop prediction models, contributing to policy making and strategic planning.

According to relevant literature the relation between statistics and machine learning consists of an increase in data complexity and the number of input variables and their possible associations make classical statistical inference less tractable and precise. While in such cases we could use ML approaches instead to fill in the unobserved aspects of the system while being effective even when the data are gathered without a carefully controlled experimental design and in the presence of complicated nonlinear interactions (Bzdok et al. 2018). In similar cases, we could also use ML to extract information from data more effectively (Zhang et al. 2022). ML tools and techniques provide means for empirical validation e.g., machine learning proved to be essential in understanding and linking indicators and indices to policy, resilience, and empirical data, contributing to a better understanding of climate resilience (Feldmeyer et al. 2020). ML tools can expand the capabilities of traditional models e.g., capture nonlinear effects which are not detected by traditional econometric models. This has been demonstrated by detecting important factors and nonlinear relationships between regional GDP per capita and Higher Education Systems indicators that have provided useful insights and suggestions for policymakers (Bertoletti et al. 2022) or to incorporate spatial, contemporaneous, and historical dependencies e.g., lead-lag non-linear relationships among past urban changes in each region and its neighbors (Kim et al. 2022). As indicated above, the discussion in literature of comparing traditional models (mainly statistical) with ML models is active. In many applications, ML models performed better than statistical models, e.g., predict particulate matter (Kulkarni et al. 2022) or suicides (Grendas et al. 2022). ML models have improved the existing models when combined with statistical ones, e.g., Alzheimer's disease (Tan et al. 2021). They have also optimized model calibration (Amroun et al. 2022). Although the core of these techniques uses mathematical models; the field of the search is significantly expanded by the acceleration and automation provided by computers.

As mentioned, data techniques used in various cases can be found in the literature, mainly concerning specific and more focused fields or topics than general ones. Until recently, related research of using ML tools in regional science focused on overall sustainability and performance, and resilience of regions had been limited, especially compared to other fields such as medical applications. In recent years, the standard has transitioned to a comprehensive and overall study of regions and areas using such tools e.g., using decision trees for regional Development Classification Models (Munandar, Winarko 2015). A recent study of resilience focused on earthquakes using historical data from previous seismic events and long-term historical behavior of regions (Fantechi, Modica 2022) is another example of combining traditional econometric with ML techniques (Bertoletti et al. 2022), which can apply ML to land-use change modeling (Kim et al. 2022). ML is also expected to play a major role in building better and modern Community Resilience Assessment tools by incorporating the use of big data, machine learning, and artificial intelligence to take care of spatio-temporal dynamism (Abdul-Rahman et al. 2021).

The paper adds to the debate on regional resilience by introducing the use of models through the utilization of Machine Learning. For this purpose, we use ML techniques to predict the performance of regions under shock, identify the more important attributes, and propose a methodological framework of how regional status can be translated into models. The paper is structured whereas the next section (Section 2) presents the methodology in detail, including the defined time period, the case study, the data sets, the variables, and the models used. Section 3 presents the analysis and the results of models' development and execution, whilst Section 4 illustrates the conclusions and future directions.

2 Methodology

2.1 The Case Study

This paper will focus on economic impact of Covid19 in USA Counties and in particular their change of unemployment rate during the first wave of pandemic and the related restrictions to examine the implementation of machine learning techniques and related ways/methodology to achieve this. The overview of the case is presented in Table 1 and it will be further analyzed in the next chapters.

The period between March and June of 2020 is termed the "Disaster Period" and defines the event studied for the selected regions in terms of their resilience and especially for this case, their ability to handle the increase of unemployment during the restrictions period. The information and related data that exist until the start of this period are considered as the current situation of the regions. These are considered as the input of the models. On the other hand, changes in values of various regional statistics during the disaster period -or values just after the end of the period- indicate how much they were affected -absolutely and comparatively- and thus are considered as performance and resilience indicators and as the output for the models.

2.2 Data Semantics and Dimensions

The current research mainly studies the response of the unemployment rate -not as unique but as a commonly accepted indicator of economic performance- in the specific disturbance defined as first wave and related restrictions of Covid19. Resistance and recovery belong to the current field of research due to both being types of regional responses. The results of the prediction models can contribute to the determination of resistance and recovery, and therefore to the degree of resilience of the areas.

In addition to the challenge itself (the efficient operation of machine learning models in regional science), great challenges are also identified in finding and properly adapting

Table 1: Case Study Details

Disaster Period: First wave of Covid19 - First Period of Restrictions and Impacts; March 2020 – June 2020 (4 months)

Disaster amplitude: Stay at home order restrictions start from March 2020 and duration up to 4 months

Input/Output

Input: Statistics regarding demographics, economy, business & industries, commuting & mobility, health, social, geographical, and other factors per county. Mainly referred in the 2019-2018 records/status and in percentages of the county's totals.

Output: Change of Unemployment Rate (March 2020 – July 2020) per County

Includes/Excludes

Includes: All the counties of USA (mainland, 3107 counties with mean population 104k)

Excludes: States of Alaska, Hawaii, and Puerto Rico (not in mainland) County Rios Aribas in New Mexico (due to data issues)

Testing Tool: Weka Platform – University of Waikato; Weka is an open-source machine learning software, widely used for teaching, research, and industrial applications (Frank et al. 2016).

available data to make the tested and applied techniques work. Appropriate input and output of the models should be clearly defined and selected with right semantics and dimensions. In this direction, the following should be defined:

- The time frame of "disaster period" for which resilience and correspondence of areas are studied: in this case, the disaster period is defined as the first Covid wave and the related stay-at-home orders of the states and in some cases of the counties (autonomously), which are generalized in the USA at the time between March and June of 2020.

- The amplitude of the disaster: as an assumption is related mainly to the duration of the restriction's orders (the longer the restriction period, the greater disaster) and secondary to their starting date (not so much concerning the disaster size, but as an extra comparison indicator for similarity between disasters of different counties). As we study the resilience and mainly the economic impact on the areas, pandemic data such as cases and deaths were considered irrelevant (or indirect factors), while the focus was on the restrictions that were raised by authorities (probably implied and forced by cases and deaths. If the resilience of the health system(s) was studied, then these parameters could be considered as direct) and have affected directly the businesses and the mobility of the counties. Other related data e.g., number of business closed or other market related parameters were strongly considered to be part of the research, but their collection was not possible due to unavailability. Therefore this is considered to be part of future research focused in regions where the related data are available.

- The areas/regions: in this case US Counties will be defined as different instances of the structure defined below.

- The data set as the set or subset of the instances (areas/regions) used to train and test the model: in this case the subsets as defined in Subsets of examination test.

- The data:
  - Appropriate input/output as attributes (values per instance) which together constitute the basic data structure:
    * input as current state: most recent stats before the start date of disaster period or constants/slow changing characteristics of the counties
    * output as performance indicators: to be the change of examined value during disaster period (or similar metrics taking into consideration seasonal adjustment).

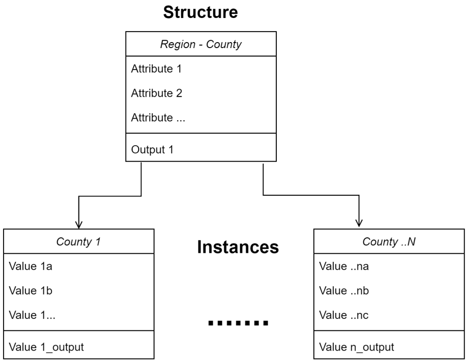
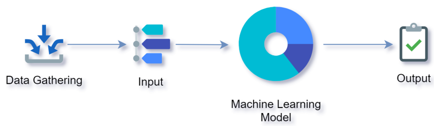


Figure 1: Structure of data



Note: graphs created in app.diagrams.net

Figure 2: Flow of the Models

  - Output attributes should be clustered properly as the models are better in predicting clusters of performance and not specific values (good or bad performance). Cluster could be applied with specific techniques e.g., K-Means or by just filtering values e.g., greater or smaller than a value.
  - Both input/output attributes should be proportional and representative meaning that date should be percentages and some other absolute numbers depending on their nature and meaning.

- Data regarding and during the "disaster period", disaster amplitude, and/or regional characteristics to find similar -in this case same restrictions due to Covid19 and/or similar population- cases/areas to define the data set required as referred above, to be studied.

Figure 1 presents the structure of data -attributes and instances- while an overall scheme of the model's flow is displayed in Figure 2.

2.3 *Subsets of examination test*

In the framework of the preliminary and main research of this study, several subsets were tested (with several criteria such as similarity or variance of population, Covid restrictions, counties in Neighboring or similar e.g., coastal states) in order to explore the application of ML models. There are unlimited sets and subsets that can be tested or demonstrated; most of which are very difficult to result in efficient prediction models. The research and comparison of different data sets and the effect of their similarities or differences in the models is included in our future research. Within the scope of this paper and based on the results of the execution, the two subsets below were selected as indicative to present our main methodology and the factors taken into consideration while translating real- life information to data sets for the purpose of machine learning techniques. To differentiate between data sets, the desire was to display that models can be created in both types of models (general or more focused with similar population category and disaster amplitude). That is why these two subsets were selected.

Subset A is a generic subset, from all the available US counties, of counties that performed "good" or "bad" during the restrictions. Subset B is more homogeneous as it includes counties that performed "good" or "bad" but also maintain a large population

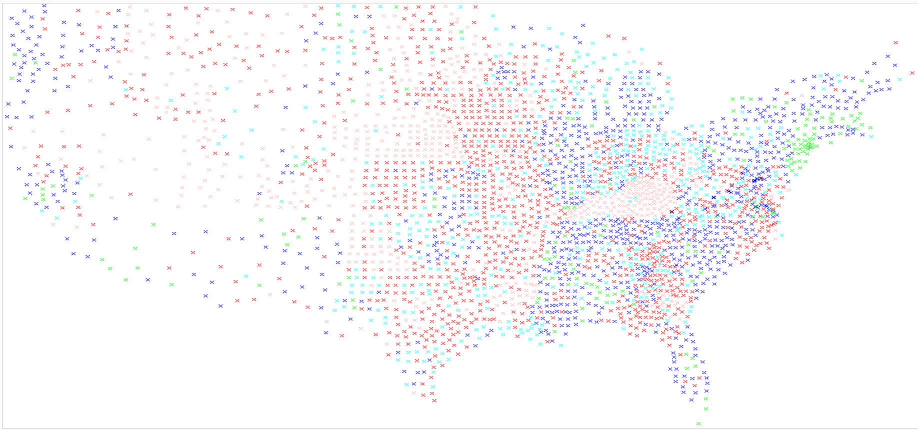


Figure 3: From WEKA: Indicative Map of US Counties

(100k and above). Counties in subset B had more similarities in the amplitude of the disaster faced (stay-at-home-order duration and when order declared).

The subsets were tested to create models predicting good or bad performance in their output, meaning the increase of unemployment rate during disaster period for the counties included.

Subset A: counties that:

- performed good (0-0,7%) or bad (7% and over increase of Unemployment Rate)
- from all available counties
- therefore 377 counties/instances

Subset B: counties that:

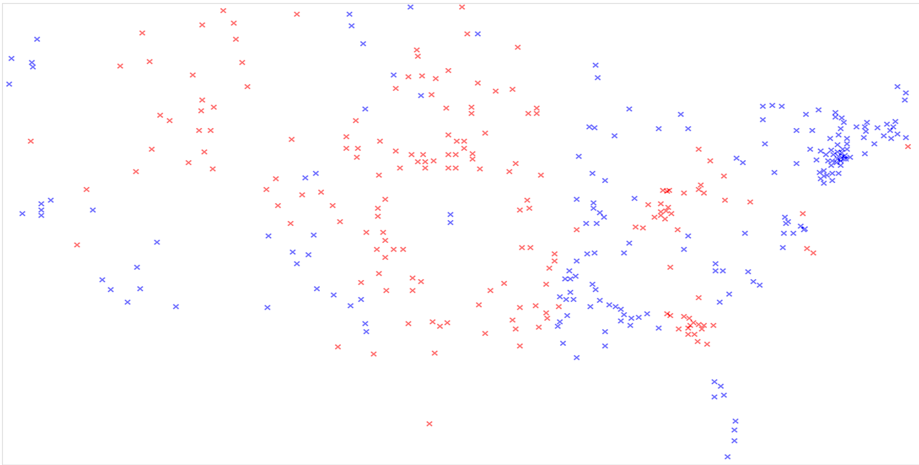
- performed good (under 2,5%) or bad (6% and over increase of Unemployment Rate)
- from counties with population > 100k and same restrictions (17-40% of the disaster period covered with stay-at-home-order and order declared soon in the first 4,5% of the period)
- therefore 89 counties/instances

Figure 3 is an indicative USA map with all the counties, Figures 4 and 5 present the specific subsets A and B of counties defined in that map.

2.4 Attributes selection and values

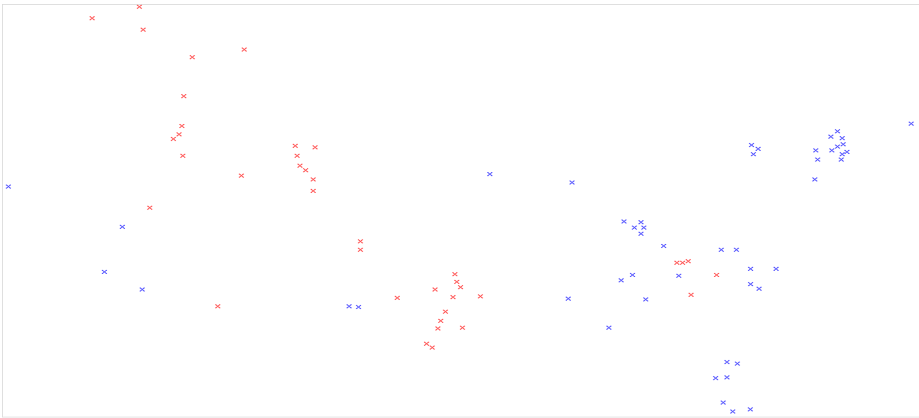
The selection of the appropriate attributes referred to in Section 2.2 is crucial to apply ML models and provide useful results. Below the selection of the three main types of data is discussed, while Appendix A displays the list of the selected attributes used in the models' execution, along with description and sources.

- Input attributes were selected based on demographic, industrial, employment, commute, and mobility, social, environmental, and health sector factors. As already mentioned, data mostly consists of percentages but sometimes absolute numbers dependent on their nature. Also, for the use of a-priori technique to find rules, data were converted from numerical into classes. The selection of the attributes and/or their type/category is generally based on regional science and literature in both fields of theory and on specific examples of indicators used in models (e.g., Feldmeyer et al. 2020, Jackson et al. 2019, Munandar, Winarko 2015). The search for useful attributes was performed in many factors affecting regional resilience. As



Note: bad in blue, good in red

Figure 4: From WEKA: Subset A – 377 Counties Performance



Note: bad in blue, good in red

Figure 5: From WEKA: Subset B – 89 Counties Performance

proposed by Christopherson et al. (2010), it is important each factor to be different in each region and some examples are a diversified economic base, the existence of competitiveness, a regional system that supports innovation and learning, partnerships, supportive financial system, modern production base which has modern infrastructure, an innovative workforce, and of course the existence of a supportive system of governance. Proposed variables estimated to relate pandemic with regional conditions are also taken into consideration (Killeen et al. 2020). Finally, any extra interesting indicator identified was selected and tested. The general approach is the constant addition of several attributes (many times even correlated with each other) to be tested and validated though the ML prediction models. ML models can be effective even when the data are gathered without a carefully controlled experimental design and in the presence of complicated nonlinear interactions (Bzdok et al. 2018).

- Output was clustered to fit prediction models which required output. The discontinuously distinct clusters (like these demonstrated here) seem to have better performance, while the prediction of the continuous ones is a much more complex and difficult problem.
- Other attributes used to cluster counties into interesting subsets to be tested (restrictions and pandemic stats) or other such as geographic coordinates to study the geographical distribution and nature of the findings.

Table 2: Performance of Models

Model	Subset A	Subset B
Multilayer Perceptron	85.15 %	91.01 %
J48	85.41 %	82.02 %
Naive Bayes	79.58 %	86.52 %

Table 3: Multilayer Perceptron Results (tested in subset B)

Model	Classified as Bad	Classified as Good
Real Class Bad	45	4
Real Class Good	4	36

2.5 Models used

In order to execute ML models, counties are defined as Instances having several attributes as Input and an Output Class. Models produced from classifiers try to classify this out class as “good” or “bad” using the input attributes. The model’s performance is the percentage of the correctly classified instances to the whole set. The specific techniques seen below were tested. The most interesting cases and these with best performance are presented in detail later in this study:

- Classifiers:
  - Multilayer Perceptron (Neural Network)
  - J48, Random Tree, REPtree (trees)
  - Naive Bayes
  - Decision Tables, JRip, OneR (rules)
  - AdaBoostM1, Attribute Selection (e.g., wrapper selecting best subsets of attributes), Stacking, Bagging (Meta Classifiers)
- A priori (association method, not a classifier, produces rules associating any input/output attribute)

3 Testing and results

3.1 Classifiers models

As discussed, many tests in different models and with different sets of input attributes were tested in the framework of the study. A performance table (Table 2) shows the three most interesting models created in both A and B subsets using selected input attributes, which show the percentage of Correctly Classified Instances using mainly the 10-fold cross-validation. This validation is considered the most valid and complete and it is used to separate the set as: 90% for training and 10% for testing being repeated 10 times so the whole set is tested as 10 independent tests. As displayed, the Multilayer Perceptron and Naive Bayes have a better performance in Subset B (which is the more focused -similar counties- approach), while J48 performs better in the more general Subset A.

Table 3, Table 4 and Table 5 present the full details regarding the best performance tests for every model.

All the models have a very good prediction ability with the Multilayer Perceptron able to correctly predict the performance of 81 out of 89 counties with only eight counties being incorrectly classified (4 as bad, 4 as Good). Although it is not clear without further analysis what factors affected the models’ decisions. Therefore, proportionally, the resilience of the counties can be a very useful prediction tool.

Table 4: J48 Results (tested in subset A)

Model	Classified as Bad	Classified as Good
Real Class Bad	174	25
Real Class Good	30	148

Table 5: Naive Bayes Results (tested in subset B)

Model	Classified as Bad	Classified as Good
Real Class Bad	43	6
Real Class Good	6	34

3.2 Decision Tree alternative use example

Although the decision tree models (J48, Random Tree, REPtree) are mainly used as classifiers (with separated training and testing sets), they can also be used in an alternative way; in tests using the whole set as the training set. This use of trees aims to find patterns and critical attributes and their specific critical values that may affect (or ways that one can understand based on variables) whether a county will have good or bad performance in the disaster period. An example presented below will be displayed and explained. It was trained with the100% of subset, while scoring 96,62% as a tree model. Performance, in the case where the whole set is also the training set, has a different meaning than the classify/prediction rate. It means that tree can find a "way of thinking" to describe, in this example, the performance of 86 counties correctly and only three incorrectly.

The main purpose of the tree is to visualize rules that result in a decision, in this case about whether a county is estimated/predicted to perform good (under 2,5%) or bad (6% and over increase of unemployment rate) during the specific disaster. A simplified visualization of the tree is displayed in Figure 7. It was produced in the Weka machine of the corresponding "code" is presented in Figure 6. We can detect, based on the output of the model, specific factors that can affect the performance of a county. It is also important that we can see specific values involved.

In this specific case, we can see that the commute time and way, vehicles available, and work from home affected the performance and the change of unemployment rate of the counties during the first wave restrictions of Covid-19. Some conclusions that can be produced from the figure are:

- If the presence of people that have a commute time above 30 minutes in their work and they drive alone (attribute: Long Commute Drives Alone) is under or equal to 17% in some county, this means that this county will perform "good".
- If the above is above 17%, but the percentage of people owning 1 vehicle is below or equal to 14,8%, then the (attribute: 1 Vehicle PCT) then by chance 90% (18.0/2.0 referred to the output of Figure 6) this county will perform "good".
- And similar for all the levels of the tree

It may be obvious that some parameters could positively affect the performance of the county (e.g., better commuting conditions will affect a lot of regional aspects including resilience), but models like this provide specific numbers e.g., the referred 17% of long commute driving alone or 0,26% commuting with bicycle. These specific numbers are an additional level of information.

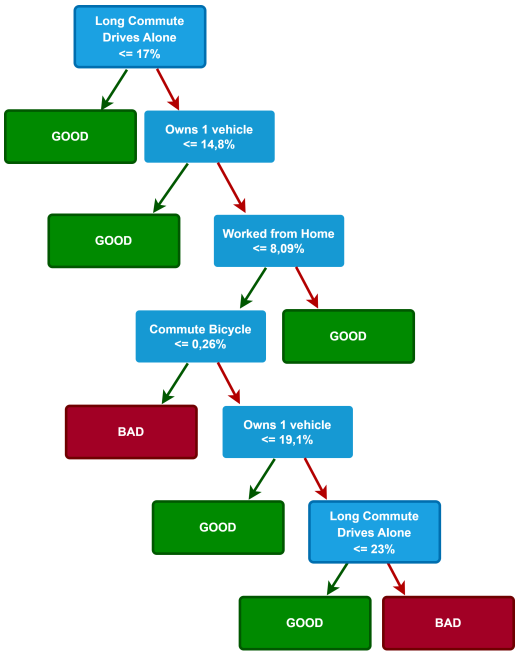
4 Discussion & Conclusion

Based on the work, practices, and approaches described in Section 2, an overall methodological framework (displayed in Figure 8) has been developed. The first step is to find

Weka output: J48 pruned tree

```
Long Commute - Drives Alone_PCT <= 17: good (12.0)
Long Commute - Drives Alone_PCT > 17
|   1_Vehicle_PCT <= 14.8: good (18.0/2.0)
|   1_Vehicle_PCT > 14.8
|   |   Worked from home PCT <= 8.09
|   |   |   Bicycle PCT <= 0.26: bad (33.0)
|   |   |   Bicycle PCT > 0.26
|   |   |   |   1_Vehicle_PCT <= 19.1: good (6.0)
|   |   |   |   1_Vehicle_PCT > 19.1
|   |   |   |   |   Long Commute - Drives Alone_PCT <= 23: good (3.0/1.0)
|   |   |   |   |   Long Commute - Drives Alone_PCT > 23: bad (13.0)
|   |   Worked from home PCT > 8.09: good (4.0)
```

Figure 6: From WEKA: Output - Tree

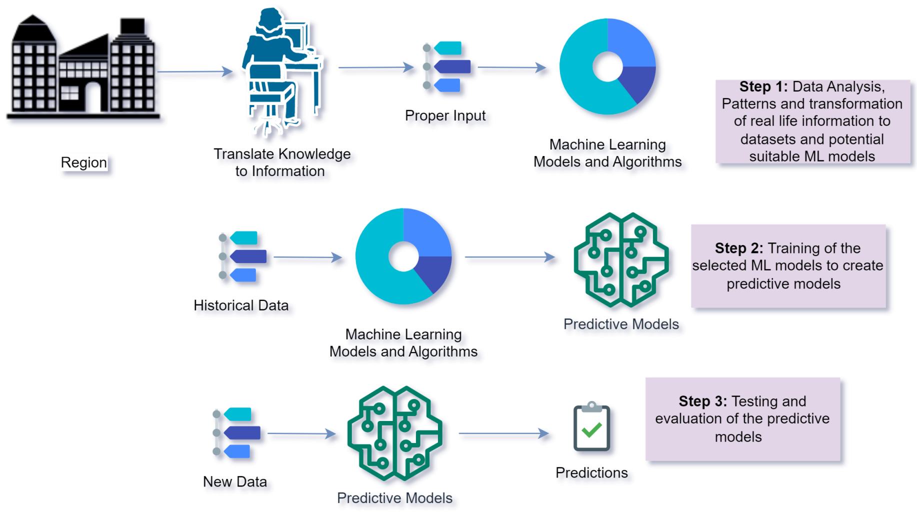


Note: graph created in app.diagrams.net

Figure 7: Decision tree

appropriate study cases (Section 2.1) and translate the regional information to appropriate input for the predictive models. Appropriate input includes forming data sets (Section 2.2) for training and testing the models, as well as the crucial step of defining the appropriate semantics and dimensions of the data (e.g., meaningful input and clustering for the output). Based on these data sets, several machine models can be selected for testing and evaluation. Historical data are used for training the ML models, resulting in some predictive models (step 2), which are then tested and evaluated according to their ability to predict the defined output on new (not used before for the training) input data set(s) (step 3).

This study has created several ML models predicting the performance of a region during disasters and has found parameters that may affect the resilience of the regions. It also has presented the above methodology and useful guidelines in how regional factors can be translated and processed by data and ML/AI tools and techniques, thus creating related models like the ones that have been demonstrated. As for the models, some like the Multilayer Perceptron and Naive Bayes seem to have a better performance in focused and homogeneous data sets, while others such as J48 have better results in general data



Note: graph created in app.diagrams.net

Figure 8: Methodology Overview

sets. Furthermore, although all models have done well in predicting the performance of the counties; decision trees offer more semantics and human readable information from the other models offering specific parameters and values that affect the results and thus may affect the resilience of an area. Multilayer Perceptron and Neural Networks function like "black boxes" and it is difficult to extract information and readable conclusions from their equations.

As for the development of these models using regional, social, economic, and environmental factors and indicators and examining regional properties related to sustainability and resilience; there are challenges on how to properly adjust and translate these real-life data and properties into appropriate data to make ML models work. A key contribution of this study is that it presents a methodology, examples, and practices on how to represent regional factors in terms of data for the input of models (as attributes), the time frame and the amplitude of disaster, and the areas/regions (as instances) forming the required data set. It indicates how to choose the appropriate input and output of the models from this data set, clearly defined and selected with right semantics and dimensions, and how to find and properly adapt the available data to make the tested and applied techniques work.

As for the models' execution and demonstration, it was displayed that creating models for predictions related to regional properties and especially for resilience that having satisfying performances is possible and deserves the attention of the regional scientists and potentially could support decisions in policy making and regional development strategies.

4.1 Future research suggestions and challenges

Data sectors, along with their increasing computing power could support regional research to expand beyond the classic math, quantitative methods, and statistical analysis, contributing to the automation in the development and validation of models though hidden paths and performing correlations and calculating combinations whose calculation using traditional methods would consume an unrealistic amount of time. The added value of ML in other fields and especially in technical issues is already examined in the literature (Section 2.2). The verification of the corresponding added value of ML and its application in the less technical field of regional development is an important field of future further research. Additionally, the appropriate selection of different model types having different performance in different cases and types of test sets (in terms of focused or general, similar, or different regions and amplitude of disasters, small or big data etc.) should be strongly considered. As far as the selection of input attributes is concerned, it was based on regional science theories. However, any extra interesting indicator that was identified was tested as ML models can handle many parameters independently of their correlation. Specifically, ML models include algorithms for the appropriate selection of sets and subsets of input attributes. They can also calculate any correlation between all input parameters (each other) and any contribution in the output. Therefore, the proposed general approach is the constant addition of several attributes (even correlated) and factors to be tested and validated though ML prediction models. This effort can be further studied to enhance the function of the models measuring the effect of the input parameters, their category, and/or their number in the model's performance. A further analysis should be dedicated to regional dimensions and direct related parameters (e.g., coordinates) by reason that coordinates were not used as input attributes in this study. Another issue that should be studied is the understandability of the ML models created. As discussed, classifiers such as Neural Networks (e.g., Multilayer Perceptron used) function as "black boxes". Thus, we must find ways to decode the models and export valuable and readable information and conclusions about the factors affecting their decisions and the regional resilience. Research in this direction may be combined with statistical analysis or other classic methods. On the other hand, decision trees display specific values and variables indicating factors and values affecting the resilience. A further study should evaluate the true impact of these identified factors. It is of great importance and a great challenge to properly apply regional analysis in all information exported by the ML models and integrate this knowledge smoothly to regional science research. The use of integrated models, combining classical with ML techniques, should and will be strongly considered in our further research. The way that information resulting from predictions can be used is very crucial and the improper use of it can lead to losses instead of benefits (either technically or socially). Technical knowledge and work cannot replace the social, humanitarian, political, and environmental dimensions. This work should be used as a tool with computational and ancillary activity. Despite the challenges, these models that are utilizing the innovations on infrastructures and computer power could enhance and modernize the toolbox of regional analysis (currently mainly based on mathematical offline methods), which could reveal new patterns and regional factors that could enable calculations that were not possible before. Additionally. "real-time" results, information, and predictions could be introduced. These tools and models could be used (or even be the baseline) in the framework of the current trend of digitalization of regions towards sustainability. They could be used to exploit data collected from IoT or crowd sensing platforms, provide related features to digital tools enabling smart and sustainable regions or cities, and therefore support decisions in policy making and regional development strategies.

References

Abdul-Rahman M, Chan E, Li X, Wong M, Xu P (2021) Big data for community resilience assessment: A critical review of selected global tools. 1345–1361. CrossRef

Ahsan MM, Siddique Z (2022) Machine learning-based heart disease diagnosis: A systematic literature review. *Artificial Intelligence in Medicine* 128: 102289. CrossRef

Amroun H, Hafid F, Mehdi A (2022) How statistical modeling and machine learning could help in the calibration of numerical simulation and fluid mechanics models? Application to the calibration of models reproducing the vibratory behavior of an overhead line conductor. *Array* 15. CrossRef

Bai Y, Zhao M, Li R, Xin P (2022) A new data mining method for time series in visual analysis of regional economy. *Information Processing & Management* 59: 102741. CrossRef

Berghout T, Benbouzid M, Muyeen SM (2022) Machine learning for cybersecurity in smart grids: A comprehensive review-based study on methods, solutions, and prospects. *International Journal of Critical Infrastructure Protection* 38: 100547. CrossRef

Bertoletti A, Berbegal-Mirabent J, Agasisti T (2022) Higher education systems and regional economic development in europe: A combined approach using econometric and machine learning methods. *Socio-Economic Planning Sciences* 82. http://www.sciencedirect.com/science/article/pii/S003801212200009X

Briguglio L, Cordina G, Bugejia S, Farrugia N (2006) Conceptualizing and measuring economic resilience. University of Malta Working Paper, 13. https://www.researchgate.net/publication/229039198_Conceptualizing_and_measuring_economic_resilience

Bristow G (2010) Resilient regions: re-'place'ing regional competitiveness. *Cambridge Journal of Regions, Economy and Society* 3: 153–167. CrossRef

Bzdok D, Altman N, Krzywinski M (2018) Statistics versus machine learning. *Nature Methods* 15. CrossRef

Christopherson S, Michie J, Tyler P (2010) Regional resilience: theoretical and empirical perspectives. *Cambridge Journal of Regions, Economy and Society* 3: 3–10

Dhar TK, Khirfan L (2017) A multi-scale and multi-dimensional framework for enhancing the resilience of urban form to climate change. *Urban Climate* 19: 72–91

Fantechi F, Modica M (2022) Learning from the past: a machine-learning approach for predicting the resilience of locked-in regions after a natural shock. *Regional Studies*. CrossRef

Feldmeyer D, Wilden D, Jamshed A, Birkmann J (2020) Regional climate resilience index: A novel multimethod comparative approach for indicator development, empirical validation and implementation. *Ecological Indicators* 119: 106861. CrossRef

Foster KA (2007) A case study approach to understanding regional resilience. Working Paper 2007-08, Institute of Urban and Regional Development, Berkeley. https://www.econstor.eu/obitstream/10419/59413/1/592535347.pdf

Frank E, Hall MA, Witten IH (2016) The weka workbench. Online appendix for "data mining: Practical machine learning tools and techniques", morgan kaufmann, fourth edition, 2016 https://www.cs.waikato.ac.nz/ml/weka/Witten_et_al_2016_appendix.pdf

Grendas LN, Chiapella L, Rodante DE, Daray FM (2022) Comparison of traditional model-based statistical methods with machine learning for the prediction of suicide behaviour. *Journal of Psychiatric Research* 145. CrossRef

Hill EW, Wial H, Wolman H (2008) Exploring regional economic resilience. Institute of Urban and Regional Development (IURD), Berkeley, Working Paper 2008-04 https://escholarship.org/uc/item/7fq4n2cv

Holling CS (1973) Resilience and stability of ecological systems. *Annual Review of Ecology and Systematics* 4: 1–23. CrossRef

Jackson RW, Hewings GJD, Rey S, Lozano-Gracio N (2019) Regional development: Challenges, methods, and models. WVU Research Repository, 2020 https://researchrepository.wvu.edu/rri-web-book/2/

Kallioras D (2011) Η Έννοια της Ανθεκτικότητας στην Περιφερειακή Επιστήμη (the concept of regional resilience in regional science). *Σειρά Ερευνητικών Εργασιών Πανεπιστημίου Θεσσαλίας* 17: 69–86. CrossRef

Killeen BD, Wu JY, Shah K, Zapaishchykova A, Nikutta P, Tamhane A, Chakraborty S, Wei J, Gao T, Thies M, Unberath M (2020) A county-level dataset for informing the United States' response to COVID-19. Technical report. CrossRef

Kim Y, Safikhani A, Tepe E (2022) Machine learning application to spatio-temporal modeling of urban growth. *Computers, Environment and Urban Systems* 94. CrossRef

Kulkarni P, Sreekanth V, Upadhya AR, Gautam HC (2022) Which model to choose? performance comparison of statistical and machine learning models in predicting PM2.5 from high-resolution satellite aerosol optical depth. *Atmospheric Environment* 282. CrossRef

Mangalathu S, Karthikeyan K, Feng DC, Jeon JS (2022) Machine-learning interpretability techniques for seismic performance assessment of infrastructure systems. *Engineering Structures* 250: 112883. CrossRef

Martin R (2012) Regional economic resilience, hysteresis and recessionary shocks. *Journal of Economic Geography* 12: 1–32. CrossRef

Mirzaei K, Arashpour M, Asadi E, Masoumi H, Bai Y, Behnood A (2022) 3d point cloud data processing with machine learning for construction and infrastructure applications: A comprehensive review. *Advanced Engineering Informatics* 51: 101501. CrossRef

Munandar TA, Winarko E (2015) Regional development classification model using decision tree approach. *International Journal of Computer Applications* 114

Patel V, Shah M (2022) Artificial intelligence and machine learning in drug discovery and development. *Intelligent Medicine* 2: 134–140. CrossRef

Paudel D, Boogaard H, Wit A, Velde M, Claverie M, Nisini L, Janssen S, Osinga S, Athanasiadis I (2022) Machine learning for regional crop yield forecasting in Europe. *Field Crops Research* 276: 108377. CrossRef

Pendall R, Foster K, Cowell M (2010) Resilience and regions: building understanding of the metaphor. *Cambridge Journal of Regions, Economy and Society* 3: 71–84. CrossRef

Pickett STA, Cadenasso ML, Grove JM (2004) Resilient cities: Meaning, models, and metaphor for integrating the ecological, socio-economic, and planning realms. *Landscape and Urban Planning* 69: 369–384

Proag V (2014) Assessing and measuring resilience. *Procedia Economics and Finance* 18: 222–229. CrossRef

Qezelbash-Chamak J, Badamchizadeh S, Eshghi K, Asadi Y (2022) A survey of machine learning in kidney disease diagnosis. *Machine Learning with Applications, Volume 10, 100418, ISSN 2666-8270*. CrossRef

Seeliger L, Turok I (2013) Towards sustainable cities: Extending resilience with insights from vulnerability and transition theory. *Sustainability* 5: 2108–2128

Serpieri C, Pontarollo N (2018) *A composite policy tool to measure territorial resilience capacity*. Publications Office of the European Union, Luxembourg. CrossRef

Shehab M, Abualigah L, Shambour Q, Abu-Hashem MA, Shambour MKY, Alsalibi AI, Gandomi AH (2022) Machine learning in medical applications: A review of state-of-the-art methods. *Computers in Biology and Medicine* 145: 105458. CrossRef

Tan MS, Cheah PL, Chin AY, Looi LM, Chang SW (2021) A review on omics-based biomarkers discovery for alzheimer's disease from the bioinformatics perspectives: Statistical approach vs machine learning approach. *Computers in Biology and Medicine* 139. CrossRef

United Nations (2016) *The World's cities in 2016*. United Nations Press, New York, NY

Zhang M, Tang X, Zhao S, Wang W, Zhao Y (2022) Statistical arbitrage with momentum using machine learning. *Procedia Computer Science* 202: 194–202. CrossRef

A Appendices

A.1 Input Attributes and other variables per County used in the research

NAME OF ATTRIBUTE	DESCRIPTION	SOURCE
POPESTIMATE2019	Estimation of population 2019 – Used mainly to clustering subsets and not as input	https://www2.census.gov
FEMALE_PCT	% of women in population	https://data.census.gov/
Race and ethnicity Percentages	% of people in specific ethnicities	https://www.countyhealth-rankings.org
NOT PROFICIENT IN ENGLISH_PCT	% of people	https://www.countyhealth-rankings.org
RDOMESTICMIG2019	Net domestic migration rate in period 7/1/2018 to 6/30/2019	https://data.census.gov/
HOUSEHOLD_AVRG_SIZE	average size of households	https://data.census.gov/
Age Groups Percentages	% of people in specific age group	https://data.census.gov/
PRIVATE_WORKERS_-PRC	% of workers in the specific type of employment	https://data.census.gov/
SELF_INCORPORATE_-WORKERS_PRC	% of workers in the specific type of employment	https://data.census.gov/
PRIVATE_NON_PROFIT_-WORKERS_PRC	% of workers in the specific type of employment	https://data.census.gov/
GOV_MUN_FEDERAL_-WORKERS_PRC	% of workers in the specific type of employment	https://data.census.gov/
SELF_NON_INCORPO_-FAMILY_WORKERS_PRC	% of workers in the specific type of employment	https://data.census.gov/
COMMUTE_TIME_X_Y_-PRC	% of people with commuting time to work x to y minutes e.g. 0-14 or 15-30 etc.	https://data.census.gov/
DRIVE ALONE TO WORK PCT	% of people driving alone to work	https://www.countyhealth-rankings.org
CAR, TRUCK, OR VAN PCT	% of people with the specific commuting way	https://data.census.gov/
PUBLIC TRANSPORTA-TION (EXCLUDING TAXICAB) PCT	% of people with the specific commuting way	https://data.census.gov/
WORKED FROM HOME PCT	% of people with the specific commuting way	https://data.census.gov/
NOVENICLE_PCT	% of people with no vehicle	https://data.census.gov/
N_VENICLE_PCT	% of people with N vehicle(s)	https://data.census.gov/
HOMEOWNERS_PCT	% of homeowners	https://www.countyhealth-rankings.org
BEST_INDUSTRY_CLASS	Class of the biggest industry (in terms of employment, of the 20 main industry sectors Appendix B – Industries Classes)	https://data.census.gov/
Industry classes percentages	Industry information in the 20 main industry sectors (e.g. proportional size of each class/sector Appendix B – Industries Classes in terms of employment to the whole employment force)	https://data.census.gov/
UNINSURED_ADULTS_-PCT	% of population under age 18-65 without health insurance	https://www.countyhealth-rankings.org

NAME OF ATTRIBUTE	DESCRIPTION	SOURCE
INCOME_INEQUALITY_-RATIO	Ratio of household income at the 80th percentile to income at the 20th percentile	https://www.countyhealth-rankings.org
SOCIAL_ASSOCIATIONS_-RATE	# of membership associations per 10,000 population	https://www.countyhealth-rankings.org
SOME_COLLEGE_PCT	% of adults ages 25-44 with some post-secondary education	https://www.countyhealth-rankings.org
HEALTH_RATE_FACTOR	A factor combining Adult smoking, Adult obesity, Food environment index, Physical inactivity, Access to exercise opportunities, Excessive drinking, Alcohol-impaired driving deaths	https://www.countyhealth-rankings.org
AIR POLLUTION - PAR-TICULATE MATTER	Average daily density of fine particulate matter in micrograms per cubic meter (PM2.5)	https://www.countyhealth-rankings.org
MEDIAN HOUSEHOLD INCOME 2019	Median Household Income 2019	https://www.countyhealth-rankings.org
ALAND	Land Area – – Used mainly to clustering subsets and not as input	https://www2.census.gov/geo https://www.census.gov/geo-graphies/
INTPTLAT	coordinate – NOT used as input, only for demonstration of results	https://www2.census.gov/geo https://www.census.gov/geo-graphies/
INTPTLONG	coordinate – NOT used as input, only for demonstration of results	https://www2.census.gov/geo https://www.census.gov/geo-graphies/
DEATHS_TILL_JUNE_PCT	% of deaths per population till JUNE 2020 – NOT used as input, only to clustering subsets of datasets – NOT used as input, only to clustering subsets of datasets in the overall research	https://usafacts.org/visuali-zations/coronavirus-covid-19-spread-map/
CASES_TILL_JUNE_PCT	% of cases per population till JUNE 2020 – NOT used as input, only to clustering subsets of datasets in the overall research	https://usafacts.org/visuali-zations/coronavirus-covid-19-spread-map/
DEATHPERCACE_PCT	% of deaths per cases till JUNE 2020 – NOT used as input, only to clustering subsets of datasets in the overall research	https://usafacts.org/visuali-zations/coronavirus-covid-19-spread-map/
STAY_AT_HOME_PCT	% of period with stay-at-home order in power (base: period 1/3/20-30/6/2020) (how much they stay in lockdown) – NOT used as input, only to clustering subsets of datasets A,B	https://en.wikipedia.org/wi-ki/U.S._state_and_local_go-vernment_responses_to_-the_COVID-19_pandemic https://www.nashp.org/gover-nors-prioritize-health-for-all/ https://www.cdc.gov/mmwr/-volumes/69/wr/mm6935a2.htm https://www.finra.org/rules-guidance/key-topics/covid-19/-shelter-in-place
STAY_AT_HOME_START	% of period with stay-at-home order in power (base: period 1/3/20-30/6/2020) (how much they stay in lockdown) – NOT used as input, only to clustering subsets of datasets A, B	https://www.countyhealth-rankings.org/explore-health-rankings/measures-data-sour-ces/2020-measures

NAME OF ATTRIBUTE	DESCRIPTION	SOURCE
DIFF_MAR_JUL_2020	To cluster the main output - % Change of unemployment rate Mar2020 to Jul2020 , NOT used as input	https://data.bls.gov/lausmap/-showMap.jsp
POS_GOOD_BAD	THE OUTPUT	Clustering in DIFF_MAR_JUL_2020 (different for any subset)

A.2 Industries Classes

Industry	Code
Agriculture, forestry, fishing and hunting	C1
Mining, quarrying, and oil and gas extraction	C2
Construction	C3
Manufacturing	C4
Wholesale trade	C5
Retail trade	C6
Transportation and warehousing	C7
Utilities	C8
Information	C9
Finance and insurance	C10
Real estate and rental and leasing	C11
Professional, scientific, and technical services	C12
Management of companies and enterprises	C13
Administrative and support and waste management services	C14
Educational services	C15
Health care and social assistance	C16
Arts, entertainment, and recreation	C17
Accommodation and food services	C18
Other services, except public administration	C19
Public administration	C20

A.3 Counties of Subsets

A.3.1 Subset A

Butler County Alabama, Dallas County Alabama, Greene County Alabama, Hale County Alabama, Lowndes County Alabama, Macon County Alabama, Mobile County Alabama, Montgomery County Alabama, Perry County Alabama, Washington County Alabama, Wilcox County Alabama, Yuma County Arizona, Jackson County Arkansas, Lincoln County Arkansas, Newton County Arkansas, Prairie County Arkansas, Searcy County Arkansas, Sevier County Arkansas, Woodruff County Arkansas, Alameda County California, Contra Costa County California, Kings County California, Los Angeles County California, Mono County California, Orange County California, Riverside County California, Sacramento County California, San Bernardino County California, San Diego County California, San Francisco County California, Siskiyou County California, Solano County California, Archuleta County Colorado, Chaffee County Colorado, Garfield County Colorado, Gilpin County Colorado, Las Animas County Colorado, Logan County Colorado, Mesa County Colorado, Ouray County Colorado, Pueblo County Colorado, Summit County Colorado, Teller County Colorado, Hartford County Connecticut, New London County Connecticut, Broward County Florida, Lake County Florida, Miami-Dade County Florida, Monroe County Florida, Orange County Florida, Osceola County Florida, Palm Beach County Florida, Polk County Florida, Appling County Georgia, Atkinson County Georgia, Bacon County Georgia, Ben Hill County Georgia, Berrien County Georgia, Charlton County Georgia, Clay County Georgia, Clinch County Georgia, Dodge County Georgia, Fannin County Georgia, Glascock County Georgia, Irwin County Georgia, Jeff Davis County Georgia, Long County Georgia, Marion County Georgia, Pulaski County Georgia, Schley County Georgia, Telfair County Georgia, Worth County Georgia, Bingham County Idaho, Bonneville County Idaho, Franklin County Idaho, Gooding County Idaho, Idaho County Idaho, Jefferson County Idaho, Jerome County Idaho, Oneida County Idaho, Shoshone County Idaho, Alexander County Illinois, Boone County Illinois, Coles County Illinois, Cook County Illinois, Franklin

County Illinois, Hardin County Illinois, Jefferson County Illinois, Macon County Illinois, Massac County Illinois, Peoria County Illinois, St Clair County Illinois, Saline County Illinois, Winnebago County Illinois, Orange County Indiana, Chickasaw County Iowa, Crawford County Iowa, Floyd County Iowa, Howard County Iowa, Lyon County Iowa, Sioux County Iowa, Cheyenne County Kansas, Hamilton County Kansas, Sedgwick County Kansas, Stanton County Kansas, Sumner County Kansas, Wichita County Kansas, Barren County Kentucky, Boone County Kentucky, Boyle County Kentucky, Campbell County Kentucky, Fayette County Kentucky, Franklin County Kentucky, Jefferson County Kentucky, Jessamine County Kentucky, Kenton County Kentucky, Knox County Kentucky, Lincoln County Kentucky, Madison County Kentucky, Marion County Kentucky, Mercer County Kentucky, Warren County Kentucky, Beauregard Parish Louisiana, De Soto Parish Louisiana, Franklin Parish Louisiana, LaSalle Parish Louisiana, Orleans Parish Louisiana, Richland Parish Louisiana, Sabine Parish Louisiana, St Helena Parish Louisiana, Union Parish Louisiana, Webster Parish Louisiana, Androscoggin County Maine, Cumberland County Maine, Oxford County Maine, Barnstable County Massachusetts, Berkshire County Massachusetts, Bristol County Massachusetts, Essex County Massachusetts, Franklin County Massachusetts, Hampden County Massachusetts, Hampshire County Massachusetts, Middlesex County Massachusetts, Nantucket County Massachusetts, Norfolk County Massachusetts, Plymouth County Massachusetts, Suffolk County Massachusetts, Worcester County Massachusetts, Calhoun County Michigan, Genesee County Michigan, Muskegon County Michigan, Wayne County Michigan, Aitkin County Minnesota, Brown County Minnesota, Kittson County Minnesota, Le Sueur County Minnesota, Mahnomen County Minnesota, Norman County Minnesota, Yellow Medicine County Minnesota, Chickasaw County Mississippi, Claiborne County Mississippi, Clay County Mississippi, Coahoma County Mississippi, Hinds County Mississippi, Holmes County Mississippi, Humphreys County Mississippi, Issaquena County Mississippi, Jefferson County Mississippi, Leflore County Mississippi, Neshoba County Mississippi, Noxubee County Mississippi, Panola County Mississippi, Quitman County Mississippi, Tunica County Mississippi, Washington County Mississippi, Camden County Missouri, Daviess County Missouri, Hickory County Missouri, Mercer County Missouri, Morgan County Missouri, Shelby County Missouri, Stoddard County Missouri, Beaverhead County Montana, Chouteau County Montana, Judith Basin County Montana, Liberty County Montana, Sweet Grass County Montana, Teton County Montana, Valley County Montana, Adams County Nebraska, Arthur County Nebraska, Buffalo County Nebraska, Cass County Nebraska, Cheyenne County Nebraska, Colfax County Nebraska, Dawes County Nebraska, Dawson County Nebraska, Dodge County Nebraska, Franklin County Nebraska, Jefferson County Nebraska, Kearney County Nebraska, Kimball County Nebraska, Lincoln County Nebraska, Madison County Nebraska, Otoe County Nebraska, Phelps County Nebraska, Red Willow County Nebraska, Richardson County Nebraska, Saunders County Nebraska, Scotts Bluff County Nebraska, Seward County Nebraska, Thayer County Nebraska, York County Nebraska, Clark County Nevada, Elko County Nevada, Eureka County Nevada, Lyon County Nevada, Atlantic County New Jersey, Bergen County New Jersey, Burlington County New Jersey, Camden County New Jersey, Cumberland County New Jersey, Essex County New Jersey, Gloucester County New Jersey, Hudson County New Jersey, Hunterdon County New Jersey, Mercer County New Jersey, Middlesex County New Jersey, Monmouth County New Jersey, Morris County New Jersey, Ocean County New Jersey, Passaic County New Jersey, Salem County New Jersey, Somerset County New Jersey, Sussex County New Jersey, Union County New Jersey, Warren County New Jersey, Bernalillo County New Mexico, Chaves County New Mexico, Grant County New Mexico, Lea County New Mexico, Lincoln County New Mexico, Sandoval County New Mexico, San Juan County New Mexico, Santa Fe County New Mexico, Taos County New Mexico, Albany County New York, Bronx County New York, Broome County New York, Chemung County New York, Dutchess County New York, Erie County New York, Fulton County New York, Greene County New York, Kings County New York, Monroe County New York, Montgomery County New York, Nassau County New York, New York County New York, Niagara County New York, Oneida County New York, Onondaga County New York, Orange County New York, Orleans County New York, Putnam County New York, Queens County New York, Richmond County New York, Rockland County New York, Schenectady County New York, Suffolk County New York, Sullivan County New York, Ulster County New York, Westchester County New York, Dare County North Carolina, Edgecombe County North Carolina, Tyrrell County North Carolina, Logan County North Dakota, McIntosh County North Dakota, McKenzie County North Dakota, Rolette County North Dakota, Stark County North Dakota, Williams County North Dakota, Adams County Ohio, Gallia County Ohio, Holmes County Ohio, Huron County Ohio, Jackson County Ohio, Monroe County Ohio, Vinton County Ohio, Cimarron County Oklahoma, Texas County Oklahoma, Clatsop County Oregon, Lincoln County Oregon, Multnomah County Oregon, Wallowa County Oregon, Allegheny County Pennsylvania, Beaver County Pennsylvania, Dauphin County Pennsylvania, Delaware County Pennsylvania, Elk County Pennsylvania, Fulton County Pennsylvania, Lehigh County Pennsylvania, Luzerne County Pennsylvania, Monroe County Pennsylvania, Philadelphia County Pennsylvania, Providence County Rhode Island, Allendale County South Carolina, Cherokee County South Carolina, Chester County South Carolina, Horry County South Carolina, Marion County South Carolina, Marlboro County South Carolina, Orangeburg County South Carolina, Union County South Carolina, Buffalo County South Dakota, Day County South Dakota, Dewey County South Dakota, Faulk County South Dakota, Hutchinson County South Dakota, Jerauld County South Dakota, Oglala Lakota County South Dakota, Potter County South Dakota, Spink County South Dakota, Stanley County South Dakota, Davidson County Tennessee, Hancock County Tennessee, Haywood County Tennessee, Madison County Tennessee, Maury County Tennessee, Sevier County Tennessee, Shelby County Tennessee, Bailey County Texas, Blanco County Texas, Bosque County Texas, Carson County Texas, Comanche County Texas, Crane County Texas, Crosby County Texas, Ector County Texas, Hamilton County Texas, Hansford County Texas, Hardeman County Texas, Hemphill County Texas, Houston County Texas, Jeff Davis County Texas, Knox County Texas, Moore County Texas, Rains County Texas, Red River County Texas, Roberts County Texas, Runnels County Texas, Starr County Texas, Swisher County Texas, Terrell County Texas, Wilbarger County Texas, Yoakum County Texas, Davis County Utah, Tooele County Utah, Utah County Utah, Highland County Virginia, Lancaster County Virginia, Emporia city Virginia, Franklin city Virginia, Hopewell city Virginia, Martinsville city Virginia, Newport News city Virginia, Norfolk city Virginia, Petersburg city Virginia, Portsmouth city Virginia, Richmond city Virginia, Clark County Washington, Pierce County Washington, Clay County West Virginia, Forest County Wisconsin, Menominee County Wisconsin, Carbon County Wyoming, Park County Wyoming

A.3.2 Subset B

Jefferson County Alabama, Mobile County Alabama, Montgomery County Alabama, Tuscaloosa County Alabama, Cochise County Arizona, Yuma County Arizona, Riverside County California, Sacramento County California, Boulder County Colorado, Douglas County Colorado, El Paso County Colorado, Jefferson County Colorado, Larimer County Colorado, Mesa County Colorado, Pueblo County Colorado, Weld County Colorado, Broward County Florida, Collier County Florida, Hillsborough County Florida, Lake County Florida, Lee County Florida, Polk County Florida, Seminole County Florida, Cherokee County Georgia, Clayton County Georgia, Columbia County Georgia, Forsyth County Georgia, Hall County Georgia, Houston County Georgia, Bonneville County Idaho, Kootenai County Idaho, Wyandotte County Kansas, Prince Georges County Maryland, Hinds County Mississippi, St Louis city Missouri, Flathead County Montana, Gallatin County Montana, Missoula County Montana, Yellowstone County Montana, Clark County Nevada, Allegheny County Pennsylvania, Beaver County Pennsylvania, Berks County Pennsylvania, Bucks County Pennsylvania, Dauphin County Pennsylvania, Delaware County Pennsylvania, Lackawanna County Pennsylvania, Lehigh County Pennsylvania, Luzerne County Pennsylvania, Monroe County Pennsylvania, Montgomery County Pennsylvania, Northampton County Pennsylvania, Washington County Pennsylvania, York County Pennsylvania, Providence County Rhode Island, Charleston County South Carolina, Dorchester County South Carolina, Horry County South Carolina, Spartanburg County South Carolina, Sumter County South Carolina, York County South Carolina, Davidson County Tennessee, Hamilton County Tennessee, Montgomery County Tennessee, Rutherford County Tennessee, Sumner County Tennessee, Wilson County Tennessee, Bell County Texas, Brazos County Texas, Collin County Texas, Comal County Texas, Ector County Texas, Ellis County Texas, Grayson County Texas, Guadalupe County Texas, McLennan County Texas, Midland County Texas, Parker County Texas, Potter County Texas, Randall County Texas, Rockwall County Texas, Smith County Texas, Taylor County Texas, Williamson County Texas, Cache County Utah, Davis County Utah, Utah County Utah, Washington County Utah, Weber County Utah