Faculty of Mathematics and Information Science

WARSAW UNIVERSITY OF TECHNOLOGY

Group Project
Documentation: part 4

Authors: *Agata Makarewicz, Jacek Wiśniewski*

Thesis title: *Application for Analysis of the Economic Growth Indexes for European Countries*

Supervisor: *Agnieszka Jastrzębska, Ph.D. Eng.*

version 1.2

10.01.2022

Table of Contents

1　Abstract

This document contains post-completion documentation for the engineering group diploma thesis entitled "Application for Analysis of the Economic Growth Indexes for European Countries". It is a continuation of the previous documents „Group Project Documentation: part 1, 2 and 3". The document is dedicated to a technical part of the project and consists of the following chapters:

- Deployment documentation - requirements, libraries, hardware resources, configuration
- Installation instruction what steps are needed to get it into production
- User's manual - how to use the system

Furthermore, the final application version (code) is an appendix to this part of the documentation.

1.1　History of changes

Date	Author	Description	Version
5.01.2021	Agata Makarewicz	Template, installation instruction	1.0
7.01.2021	Agata Makarewicz	Deployment documentation added	1.1
10.01.2021	Agata Makarewicz, Jacek Wiśniewski	Rest of the chapters added	1.2

2　Vocabulary

Homepage - a webpage presented after turning on the application. It will have all of the functionalities like filtering data and generating the report.

"Read about the project" page – a webpage that will present all of the information about the project, authors and contact email addresses.

Report – content from homepage consisting of charts and results of clustering algorithms with comments.

Clustering - the task of dividing a set of objects into several groups called clusters in such a way that objects within the same cluster are more similar to each other than to objects in other clusters.

Model – machine learning algorithm used for clustering.

3   Errata to Documentation Part 1 & 2

Documentation Part 1 & 2 stated that the results of the work will be presented in the form of an application with two web pages, however third web page has been created in the process. It contains evaluation of the clustering algorithms used, depending on the values of their parameters, as well as interesting results and insights gained while working on the project.

4   Deployment documentation

Instruction for proper environment configuration presented below applies to working on Windows operating system. Approximately 8-9GB free disk space is required.

1) Install Python 3.8.2: https://www.python.org/downloads/release/python-382/ (newer releases of Python 3.8 should also work, however it was not tested)
   - During installation, check the box "Add Python 3.8 to path"
2) Install R: https://cran.r-project.org/bin/windows/base (newest version 4.1.2 or any other not older than 3.6 works)
   - Follow default installation. After completion, open R (C:\Program Files\R\R-<version>\bin\R.exe) and run following command: *dir.create(Sys.getenv("R_LIBS_USER"), recursive = TRUE).* This will create a user library for R located in C:\Users\<user>\Documents\R\win-library\<version> .
3) Open "Edit system environment variables", go to "Environment variables" and add new user variable.
   - Name: R_LIBS_USER
   - Value: C:\Users\<user>\Documents\R\win-library\<version> (path to R user library)
4) Install Git: https://git-scm.com/download/win . Follow default installation.
5) Install Microsoft Visual C++ Build Tools: https://visualstudio.microsoft.com/pl/visual-cpp-build-tools
   - Run the installer. Select only first component ("Desktop development with C++") and then proceed with the default selection of checkboxes on the right. There should be no more than 7GB of data to download. At the end of the installation it is required to restart the system.
6) Install Google Chrome or Microsoft Edge browser.
7) Create Python virtual environment. Open Command Line, enter the desired location and run following command: *python -m venv <environment_name>.* This will create a directory with separate Python executable, to which all required packages will be installed.

5   Installation instruction

1) Clone GitHub repository. Open Git Bash, navigate to the desired location and run following command: *git clone https://github.com/amakarewicz/BEngThesis*
   OR
   Download GitHub repository. Go to https://github.com/amakarewicz/BEngThesis -> Code -> Download ZIP. Extract the contents into desired directory.
2) Run Command Line (best as an Administrator) and enter the directory with the repository (BEngThesis if it is cloned / BEngThesis-main if downloaded).
   - Example: *C:\Users\agama\Documents\BEngThesis*
3) Run install_requirements.sh file, adding path to created previously Python virtual environment directory and path to R directory as command arguments.

- Command template: *install_requirements.sh "{/path/to/python/env }" "{/path/to/R}"*
- Example: *install_requirements.sh "C:\Users\agama\Documents\BEngThesis\django\bengthesis" "C:\Program Files\R\R-4.0.0"*

4) Run start_app.sh file, adding path to created previously Python virtual environment directory as command argument (same as in the previous step).
- Command template: *start_app.sh "{/path to python/env}"*
- Example: *start_app.sh "C:\Users\agama\Documents\BEngThesis\django\bengthesis"*
- If following warnings appear, do not worry, application is running.
  - *UserWarning: h5py not installed, hdf5 features will not be supported.*
  - *UserWarning: R is not initialized by the main thread. Its taking over SIGINT cannot be reversed here, and as a consequence the embedded R cannot be interrupted with Ctrl-C. Consider (re)setting the signal handler of your choice from the main thread.*

5) Run http://127.0.0.1:8000/homepage in your browser.

6 Acceptance tests

According to user's stories defined in the first part of documentation, in order to test the previously defined functionalities of the application, following tests were run.

- "Read about project" page
  1) Related user stories:
     a. *As a data scientist, I want to read about models used in the project, so that I can develop my data science skills.*
     b. *As an economist, I want to read about models used in the project, so that I can gain intuition about how they work.*
     c. *As a data scientist, I want to read about economic indicators, so that I can understand what they mean.*
  2) Conducted test:
     a. There is a "read about project" button located on the webpage. Clicking it navigates user to the web page, on which two scrollable tables are located, containing information about indicators, models, and its parameters.
  3) Result: passed

- "Home" page
  4) Related user stories:
     a. *As a data scientist, I want to choose a model, so that I can compare the results of different clustering algorithms.*
     b. *As a data scientist, I want to choose the model's parameters, so that I can verify different hypotheses (groupings).*
     c. *As an economist, I want to choose economic indicators, so that I can select the most accurate ones for the identification of the business cycle.*
  5) Conducted test:
     a. On the left side of home page there is a panel located, containing multiple widgets (dropdown lists, sliders etc.). They enable user to choose indicators and algorithm to use, and set its parameters. Widgets related to parameters show up accordingly to model selection. It is also possible to leave default settings. Clicking the "Perform clustering" button at the bottom of the panel executes clustering and results in creating a report containing a map with clustering results, model evaluation, clusters statistics and time series charts.
  6) Result: passed

- All pages
  7) Related user stories:

a. *As an economist, I want to interact with charts, so that I can get more detailed information about a particular group of countries/country.*
8) Conducted test:
    a. Charts generated on both home and report page are interactive, enabling the user to perform zooming, hovering, and more to display preferred information.
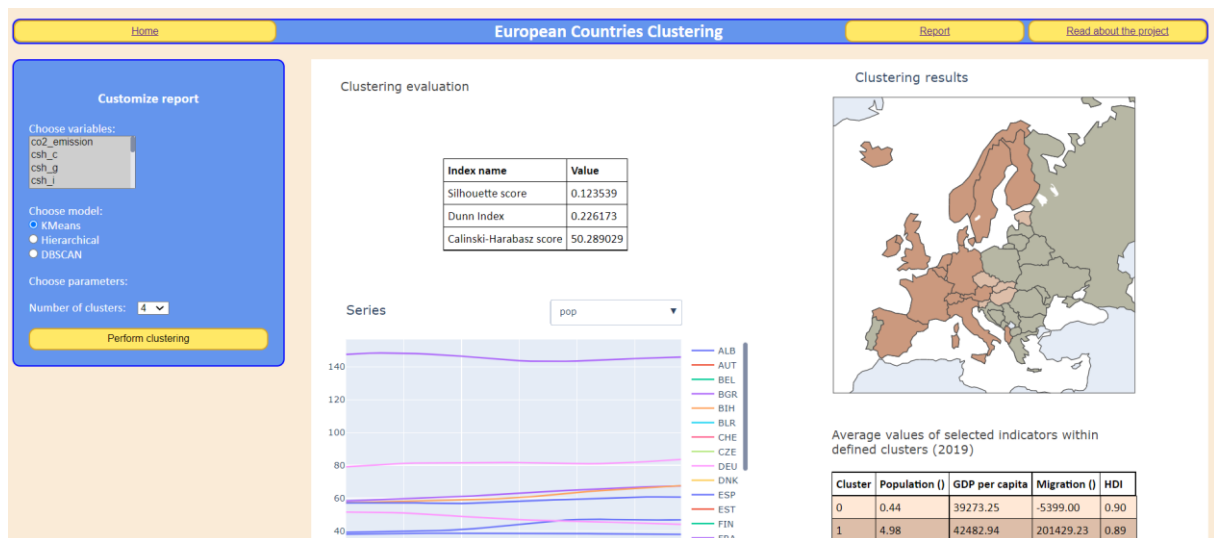9) Result: passed

7   User's Manual

Figure 1: Home page layout.

To navigate between pages, use buttons in the top right corner of the page.
- To get information about the project, datasets and algorithms used, click "Read about project" button.
- To view evaluation of the created models and interesting results and insights, click "Report" button.
- To go back to home page, click "Home" button.

Home page:

Use panel on the left to customize the report.
- To select multiple indicators, hold Ctrl button. All indicators are selected by default.
- Use radio buttons to select clustering algorithm. For each algorithm appropriate widgets are shown, enabling parameters choice. Use them to set parameters or leave the default ones.
- To generate results, click "Perform clustering" button.

To isolate only one country on the line chart, double click on the legend next to the trace assigned to it. Click on the trace in the legend to select/deselect given country. Use a drop-down list to select indicator. Hover over the trace to see more information. Use the cursor to select and zoom desired part of the chart.
Hover over the country on the map to see more information.

Report page:

Use a drop-down list to select a metric according to which models were evaluated. Use rest of the functionalities of the line chart described above.
Use a slider to manipulate parameters values / select desired time range.
Hover over the country on the maps to see more information.

8 Bibliography

[1] Aghabozorgi, Saeed, Shirkhorshidi, Ali S., and Wah, Teh Y. Time-series clustering – A decade review. Information Systems 53 16-38, 2015.

[2] Gräbner, C., Heimberger, P., Kapeller, J., and Schütz B. Structural change in times of increasing openness: assessing path dependency in European economic integration. Journal of Evolutionary Economics 30, 1467–1495, 2020.

[3] Bartlett, W. and Prica, I. Interdependence between Core and Peripheries of the European Economy: Secular Stagnation and Growth in the Western Balkans. LSE'Europe in Question' Discussion Paper Series, LEQS Paper No. 104/2016, 2016.

[4] Hamilton, James Douglas Time Series Analysis. Princeton University Press, 1994.

[5] Pal, Avishek, Prakash, PKS. Practical Time Series Analysis. Master Time Series Data Processing Visualization and Modelling Using Python. Packt, 2017

[6] Fu-Lai Chung, Tak-Chung Fu, V. Ng and R. W. P. Luk, "An evolutionary approach to pattern-based time series segmentation," in *IEEE Transactions on Evolutionary Computation*, vol. 8, no. 5, pp. 471-489, Oct. 2004.

[7] Scikit-learn developers (BSD License), Scikit-learn User Guide: 2.3 Clustering, 2007-2021, https://scikit-learn.org/stable/modules/clustering.html

[8] Jeremy Zhang, Dynamic Time Warping. Explanation and code. 2020, Feb 1 https://towardsdatascience.com/dynamic-time-warping-3933f25fcdd

[9] Xiaoji Wan, Hailin Li, Liping Zhang, Yenchun Jim Wu, "Multivariate Time Series Data Clustering Method Based on Dynamic Time Warping and Affinity Propagation", Wireless Communications and Mobile Computing, vol. 2021, Article ID 9915315, 8 pages, 2021. https://doi.org/10.1155/2021/9915315

[10]   Siebert, J.; Groß, J.; Schroth,C. A Systematic Review of Python Packages for Time Series Analysis.Eng. Proc. 2021, 5, 22. https://doi.org/10.3390/engproc2021005022

[11]   Julio-Omar Palacio-Nino, Fernando Berzal Evaluation Metrics for Unsupervised Learning Algorithms