

Group Project  
Documentation: part 2

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 1.0

30.10.2021

Table of Contents

[1 Abstract 3](#__RefHeading___Toc2796_2030408508)

[1.1 History of changes 3](#__RefHeading___Toc800_779315115)

[2 Vocabulary 4](#__RefHeading___Toc802_779315115)

[3 Specification 5](#__RefHeading___Toc6747_630044125)

[3.1 Executive summary 5](#__RefHeading___Toc6749_630044125)

[3.2 Functional requirements 5](#__RefHeading___Toc6751_630044125)

[3.3 Non-functional requirements 7](#__RefHeading___Toc6753_630044125)

[4 Project schedule 8](#__RefHeading___Toc6755_630044125)

[5 Risk Analysis 10](#__RefHeading___Toc6757_630044125)

[6 Bibliography 11](#__RefHeading___Toc6793_630044125)

# Abstract

This document contains general design specification for the engineering group diploma thesis entitled “Application for Analysis of the Economic Growth Indexes for European Countries”. It consists of following parts:

* TODO
* TODO
* TODO
* TODO
* TODO
* TODO

The main goal of the project is to apply several standard clustering methods such as k-means, hierarchical clustering and the fuzzy c-means method to time series of economic growth to group European countries and verify the previously proposed divisions, based on different criteria such as GDP per capita, the level of industrialization or HDI. The algorithms will be evaluated using the existing cluster analysis assessment indexes, e.g. inertia, silhouette score, GAP statistic and PBM index. The thesis will be based on publicly available data, including the Penn World Table. The selection of variables itself is one of the tasks. The analysis will cover complete time series and selected segments (e.g. before and after 2008 - the year of the last financial crisis). Another issue to examine will be the aspect of similarity of time series in the context of the assessment of synchronization or non-synchronization of business cycles of selected groups of countries before and after the crisis. The implemented models will be part of the web application in which the user will be able to compare the results of the methods used, select variables and parameters for the models, as well as the development indicators presented in the charts. Visualizations of the clusters obtained with different clustering methods will also be available.

## History of changes

|  |  |  |  |
| --- | --- | --- | --- |
| **Date** | **Author** | **Description** | **Version** |
| 30.10.2021 | Agata Makarewicz | First version | 1.0 |
|  |  |  | 1.1 |
|  |  |  | 1.2 |
|  |  |  | 1.3 |

# 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** - is the task of grouping a set of objects in such a way that objects in the same group (called a cluster) are more similar to each other than to those in other groups (clusters).

**Business cycle** - intervals of expansion followed by a recession in economic activity. Fluctuations are usually characterized by general upswings and downturns in a span of macroeconomic variables.

**Segmentation** – i.e. **time-series segmentation** is a method of [time-series analysis](https://en.wikipedia.org/wiki/Time_series#Analysis) in which an input time-series is divided into a sequence of discrete segments in order to reveal the underlying properties of its source.

**Model** – machine learning algorithm used for clustering.

**Risk analysis** - the science of risks and their probability and evaluation. Probabilistic risk assessment is one analysis strategy usually employed in science and engineering.

**Schedule**  - a basic time-management tool, consists of a list of times at which possible tasks, events, or actions are intended to take place, or of a sequence of events in the chronological order in which such things are intended to take place.

**Vocabulary** - a set of familiar words within a person's language. A vocabulary, usually developed with age, serves as a useful and fundamental tool for communication and acquiring knowledge

# Specification

## Executive summary

The aim of the diploma thesis is to apply various clustering algorithms to the time series of economic growth of European countries.

## Functional requirements

## Non-functional requirements

# Project schedule

Use case

*Figure 2. Project schedule.*

# Risk Analysis



|  |  |  |
| --- | --- | --- |
| 1. **SWOT** | 1. Helpful | 1. Harmful |
|  |  |  |
|  |  |  |

# 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