UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

Faculty of Engineering, Built Environment and IT

Department of Computer Science

MIT C (Big Data Science)

MIT 808

Assignment 2 - Requirements Definition Document

31 May 2019

Marea Gwyneth Sing

u10500449

Mentor: Dr Conrad Beyers **Signature:**

Co-mentor: Mrs Anna Bosman **Signature:**

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

Faculty of Engineering, Built Environment and IT

Department of Computer Science

MIT C (Big Data Science)

MIT 808

Assignment 2 - Requirements Definition Document

31 May 2019

Marea Gwyneth Sing

u10500449

Mentor: Dr Conrad Beyers **Signature:**

Co-mentor: Mrs Anna Bosman **Signature:**

Identifying and Analysing Patterns of Financial Crises Using Self-Organising Maps

Contents

1 Project Motivation

Systemic financial crises are detrimental to the economic and social well-being of not only a nation's citizens, but also the citizens of those countries that share strong macro-financial ties with the originating crisis country. Timeous intervention by national governments or central banks are necessary to mitigate some of these costs, however this requires the development of an early warning system (EWS). While the economic literature on this topic is fairly deep, the use of machine-learning to address this type of "policy prediction" problem is still scarce.

This research project focuses on using self-organising maps (SOMs) to analyse financial crises in the European union over the period 1960 to 2016. In particular, the project will help to identify the variables that will be used in building early-warning systems, and provide a first-level understanding of the relationship between them.

The project's problem statement, goals, scope and objectives, as outlined in the initial project proposal document, are reproduced below for information purposes. The reader is referred to the proposal for a discussion on the project's assumptions and limitations.

1.1 Problem Statement

In general, the variables chosen to develop early-warning systems for financial crises are based on economic theory and known transmission mechanisms. However, given the complex, non-linear nature of these systemic events, and the importance of early policy intervention, it is not well-established whether (1) this approach captures the full set of variables that could be used to improve the accuracy, precision, and recall of an EWS, and (2) this set of variables, or perhaps a subset thereof, are consistent across the considered countries and time periods.

1.2 Goals, Scope and Objectives

The goal of this project is to produce a rigorous data analysis report, using newer machine-learning techniques, to identify the variables that appear to signal imminent financial crises, or the nature of ongoing crises, and how these variables may change across countries and/or time.

This is in support of a larger research effort to establish the statistical superiority (or lack thereof) of a variety of methodological approaches to building an EWS, ranging from extant statistical approaches to newer machine-learning approaches.

Specifically, the scope of this project is two-fold, utilizing self-organising maps (SOMs) to:

1. Identify per country, which variables appear to indicate periods of imminent and ongoing financial crisis. (This analysis will employ a supervised SOM.)

2. Identify across time, which of the countries are most similar, and analyse the cross-border transmission of financial crises (that is, whether they cluster together on the map when a crisis occurs in any one of them). (This analysis will use an unsupervised SOM.)

In addition, this will produce a unique visual representation of financial crises in two-dimensional space that might assist in furthering the understanding of intra- and cross-border transmission mechanisms.

A final objective of this project is to conduct this analysis such that it can be easily monitored and edited by supervisors and co-authors, as part of a larger effort to create high-quality reproducible research. In future, this analysis may be shared publicly as a stand-alone, interactive data analysis application.

2 User Requirements

There are currently only three users for this particular project, one student and two supervisors:

- The student should be able to provide regular updates on project progress, which the supervisors should be able to monitor and provide feedback on.

- Both the students and supervisors require a well-developed version control system, such that current and previous model iterations and data versions can be easily identified and accessed.

Should this project be extended in future to become a publicly available resource, user requirements will additionally include the ability to replicate the existing analysis and improve or modify the currently employed methodology. That is:

- An external researcher should have access to clear and precise documentation on the existing project.

- An external researcher should be able to clone the project and its associated repositories (data and code).

All users require a living document to be produced that concisely describes the methodology used, and clearly presents and analyses the SOMs produced.

3 System Requirements

This section describes the functional and non-functional system requirements necessary to address the user requirements outlined in the previous section.

Given that this is a research project, the system is required to provide a collaborative space in which current researchers, and potential future researchers can work from. In order to ensure the practicality and efficiency of this space, it needs to be well-structured and easily accessible. In particular, it needs to be well-structured in terms of version control for both the data used (time-series macroeconomic data will be updated and revised) and the model produced (for example, the optimal parameter-setting may change).

3.1 Functional Requirements

- The system shall have a separate project management space, where progress can be documented and discussed.
- The system shall keep a record of modifications to models, data, and documents.
- The system shall allow for off-line modifications which can then be synced to the online version of the project.
- In future, the system will need to allow for external researchers to download the project and/or suggest modifications to the model specifications and methodology used.
- The system shall produce a collection of SOMs: (1) one for each time period and all countries, and (2) one for each country for the entire time period:

  (1) The unsupervised SOM:
  This SOM shall be estimated using a collection of World Bank indicators[1] for each country, creating one SOM for each quarterly period from 1970 to 2016. This SOM will show the distribution of countries across the two-dimensional map for each period of time - highlighting how similar these countries are based on their macroeconomic indicators. Comparing these maps across time is also expected to show how countries cluster closer together on the map in periods of crisis when these systemic risks spill over national borders.

[1] A full list of these indicators is tabulated in the project proposal document.

(2) The supervised SOM:

This SOM shall be estimated using the same collection of World Bank indicators for each country, together with the ECB/ESRB EU crises database (crisis-indicators)[2]. This SOM will be produced for each country, showing the distribution of each country's financial stability state over time. The underlying component/feature planes (each one representing the distribution of one macroeconomic variable) is expected to show which variables may indicate periods of crisis or imminent crisis.

- The system shall evaluate the fit of the SOMs using quantitative measures, such as distortion (unsupervised SOM), accuracy, precision, and recall (supervised SOM).
- The system shall use a random forest model on the same datasets employed in the supervised SOM to extract variable importance measures to supplement SOM findings.
- The system shall produce an analytical report comprised of the relevant SOMs produced[3], as well as the methodology used to produce them.

3.2 Non-functional Requirements

- The system shall split the dataset into mutually exclusive subsets in order to tune the SOM and random forest hyperparameters, such as the size of the neighbourhood and the learning rate, or the depth of the tree and size of the forest, respectively.
- The code shall be written in R, in order for it to be easily accessible and executable (i.e. open source software).
- The report shall only reference existing results (i.e. be compiled separately from the model estimation phase) to ensure that the report can be reproduced and edited quickly.
- Previous data and models shall be clearly identified and stored in the same repository as the rest of the project.
- The code, documentation and report(s) shall be accessible from all modern browsers.

[2]A full description of these datasets can be found in the project proposal document.

[3]Relevant SOMS are those covering transitional periods, from tranquil to crisis states. The full collection of SOMs will be saved in a separate document.

4 Change Management

As outlined in the project proposal, given the nature of a research project, the problem statement and scope of this project is unlikely to change. Additionally, the user and system requirements outlined in this document relate more to how the research is to be executed and shared, rather than to the project's problem statement and scope.

However, a minor change to the final project deliverables has allowed for the report that will be produced to be a living (work-in-progress) document, that will be finalised for inclusion in the broader thesis at a later date.

References

[1] How to approach functional specification documentation - justinmind. https://www.justinmind.com/blog/functional-specification-documentation-quick-guide-to-making-your-own/, July 2018. (Accessed on 05/26/2019).

[2] Laura Brandenburg. What goes into a functional specification? https://www.bridging-the-gap.com/functional-specification/. (Accessed on 05/26/2019).

[3] Dr Prittish Dala. Mit 808 – big data science project, 2019.

[4] Ulf Eriksson. Functional requirements vs non functional requirements. https://reqtest.com/requirements-blog/functional-vs-non-functional-requirements/, April 2012. (Accessed on 05/26/2019).

[5] Ulf Eriksson. One requirements document template to rule them all — reqtest. https://reqtest.com/requirements-blog/requirements-document-template/, December 2016. (Accessed on 05/26/2019).