

REQUIREMENTS ANALYSIS DOCUMENT

The Data Mining Model for Factors of Alzheimer's and Depression

by Elif AKAR- 217CS2012

<Submission Date: 01.04.2022>

Prepared for
COMP4902 Graduation Design Project



IŞIK UNIVERSITY
COMPUTER
SCIENCE AND
ENGINEERING

Table of Contents

1. INTRODUCTION.....	3
1.1. PURPOSE OF THE SYSTEM.....	3
1.2. SCOPE OF THE SYSTEM.....	4
1.3. OBJECTIVES AND SUCCESS CRITERIA OF THE PROJECT.....	4
1.4. DEFINITIONS, ACRONYMS, AND ABBREVIATIONS.....	5
1.5. OVERVIEW	5
2. CURRENT SYSTEM.....	6
3. PROPOSED SYSTEM.....	6
3.1. OVERVIEW	6
3.2. FUNCTIONAL REQUIREMENTS	7
3.3. NON-FUNCTIONAL REQUIREMENTS	8
3.3.1. Usability.....	8
3.3.2. Reliability.....	8
3.3.3. Performance	8
3.3.4. Supportability	8
3.3.5. Implementation	8
3.3.6. Interface.....	8
3.3.7. Packaging	8
3.3.8. Legal.....	8
3.4. SYSTEM MODELS	9
3.4.1. Scenarios	9
3.4.2. Use case model.....	11
3.4.3. Object model.....	15
3.4.4. Dynamic model.....	16
3.4.5. User interface—navigational paths and screen mock-ups	18
4. GLOSSARY	20
5. REFERENCES	21

REQUIREMENTS ANALYSIS DOCUMENT

1. Introduction

Alzheimer's Disease (AD) is neurodegeneration defined as a severe deterioration in the person's mental, physical, and behavioural functions due to disorders in the function of the brain. AD is one of the most common causes of dementia. According to research conducted by World Health Organization (WHO), in 2018, approximately 55 million people had dementia [1]. People with Alzheimer's Disease account for 60-70% of dementia cases [1]. According to WHO's research, it is estimated that dementia cases will double every 20 years, reaching 65.7 million in 2030 and 115.4 million in 2050 [1].

Depression is a common severe mood disorder. Depression is characterised by moods such as sadness, feelings of guilt, low self-esteem, insomnia and loss of appetite, fatigue, and poor concentration that will affect our daily lives. It can also evolve into situations such as having physical pains and complaints without an apparent physical cause. In the most severe cases, it can lead to suicide. According to research conducted by the WHO in 2015, the total number of people with depression exceeds 300 million [2]. Depression is also the most significant contributor to nearly 800.000 million suicide deaths per year [2].

In this context, in the Data Mining Model for Factors of Alzheimer's and Depression Project, I will create a user interface in which I will present the analysis study with modelling, the testing of the hypotheses proposed at the beginning of the project, the forward predictions based on the factors, the relationship between Alzheimer's Disease and Depression, to the user through data mining, using datasets obtained as a result of research on the factors that may cause Alzheimer's Disease and Depression and also I will present a model for users to define AD and Depression using datasets that they can load into the model by using The Data Mining Model for Factors of Alzheimer's and Depression Project.

1.1. Purpose of the System

The primary purpose of the system is to use data mining to analyse with modellings the factors that may cause Alzheimer's Disease and Depression on a country basis by using datasets obtained as a result of the research, to investigate factor correlations, to test the hypothesis proposed at the beginning of the project, to make predictions related with AD and Depression, to establish a relationship between Alzheimer's Disease and Depression, and to be able to define a possible Alzheimer's Disease and Depression with the factors analysed. In addition, the

secondary purpose of the system is to develop a user interface where the model can share the analyses with modelling, hypotheses, predictions, and relationships obtained in line with the data-mining study. Thus, the model will pave the way for new analyses on the factors for Alzheimer's Disease and Depression by using the user interface.

1.2. Scope of the System

The Data Mining Model for Factors of Alzheimer's and Depression is a web-based data mining project prepared by creating a user interface that users can examine and visually observe the analysis of related factors for AD and Depression with modelling, hypotheses proposed at the beginning of the project, predictions, and the relationship between AD and Depression using datasets obtained as a result of the research. Factors for Alzheimer's Disease and Depression will be analysed using country-based datasets. According to the analysis, factors that can define Alzheimer's Disease and Depression will be obtained, and forward-looking predictions will be made. The hypotheses proposed at the beginning of the project will be proven or invalidated due to analyses and tests. The result obtained by data mining studies will be presented to the user as a user interface so that the user can investigate and analyse the model. The user interface that will be created also allows the user to load datasets to examine whether datasets are related to AD and Depression. Thus, the analysis, hypothesis tests, and predictions made from the model using these datasets are presented to the user and guide the user to understand whether they can define AD and Depression with the datasets loaded into the model.

1.3. Objectives and Success Criteria of the Project

In this project, our objectives are:

- To determine the most critical factors that cause Alzheimer's Disease and Depression due to the research and identify the datasets based on countries based on these factors.
- Propose at least five hypotheses regarding the factors that cause Alzheimer's Disease and Depression using the research.
- To analyse the factors for Alzheimer's Disease using the datasets obtained in line with the determining factors.
- To analyse the factors for Depression using the datasets obtained in line with the determining factors.

- To establish a relationship between Alzheimer’s Disease and Depression based on the analyses made and the datasets obtained.
- Make advanced predictions about Alzheimer’s Disease and Depression using analysis and datasets.
- Create a user interface where users can review, visually observe, and save analysis studies, relationships, hypotheses, and predictions using data mining.

And success criteria of the project are:

- Collect and merge datasets obtained from at least five factors with Alzheimer’s Disease and perform data analysis using these datasets.
- Collect and merge datasets obtained from at least five factors with Depression and perform data analysis using these datasets.
- Identify a relationship between Alzheimer’s Disease and Depression using datasets obtained from factors.
- Present the validity or invalidity of hypotheses by testing the proposed hypotheses and establishing a relationship between hypotheses and research.
- Present analysis studies, relationships, hypotheses, and predictions made with datasets obtained from factors using reviews and modellings in the created user interface.
- Present the model that the user can load datasets, examine the results of analyses using loaded datasets, make predictions, and test their proposed hypotheses in the created user interface.

1.4. Definitions, Acronyms, and Abbreviations

AD – Alzheimer’s Disease

WHO – World Health Organization

1.5. Overview

This documentation includes the purpose (in section 1), scope (in section 1), objectives and success criteria (in section 1) of the project called The Data Mining Model for Factors of Alzheimer's and Depression, the current systems that the project takes as an example (in section 2).) is documentation that will explain the project (in section 3), the project's functional and non-functional requirements (in section 3), which are intended to be built based on these current systems. The documentation also includes scenarios (in section 3), use-cases (in section 3) and diagrams (in section 3) that can describe the project.

2. Current System

The subject of AD and Depression has been the subject of many studies before, and the results have been presented to users using various articles or web-based user interfaces. Mental Health, produced by the Institute for Health Metrics and Evaluation, prepared to make the latest estimates of mental health disorder prevalence and associated disease burden, and reported in the Global Burden of Disease study, is a web-based and research-based up-to-date system that includes AD and Depression [3]. Various mental health diseases, including AD and Depression, were analysed in the system with their related factors, and these studies were modelled (tables, graphs, maps) and presented to the user. The user can make filters using the models with parameter changes allowed by the system [3]. Depression Rates by Country 2022, which was prepared using the WHO's Depression and Other Common Mental Disorders: Global Estimates research study as a source, can also be defined as the current system for data mining research on Depression [4]. The information that depression rates are pretty high in less developed countries, which is one of the most critical data obtained due to the research, is presented to the user with tables and models (tables, graphs, maps) [4]. Users can also observe the studies on parameters such as population, religion, and economy made on a country basis to examine their relationship with depression [4]. Along with these two web-based studies, we can define the article Risk Factors and Identifiers for Alzheimer's Disease: A Data Mining Analysis as the current system [5]. The study offers the user the outputs obtained from analysing factors such as social culture, economy, education, age, and gender that may affect AD [5]. WHO's Coronavirus (COVID-19) Dashboard is also defined as the current system [6]. It is an essential web-based resource for The Data Mining Model for Factors of Alzheimer's and Depression [6]. However, it is not related to AD or Depression. In this system, where the coronavirus cases are processed up-to-date, users can observe the current issues of countries and continents [6]. The coronavirus case processes from a specific date range to the present through tables and maps [6].

3. Proposed System

3.1. Overview

The Data Mining Model for Factors of Alzheimer's and Depression is designed as a research-based and web-based data mining project. The first part of The Data Mining Model for Factors of Alzheimer's and Depression will develop based on the datasets obtained based on countries of at least the five most critical factors for AD and Depression, determined by literature research. Analyses will be made using the datasets obtained from factors and AD and Depression rates datasets on a country basis. Analyses will be strengthened numerically with tables and visually with graphics, which means that modellings will support analyses. Thus, whether the determining factors are related to AD and Depression will be defined. After the analyses, future predictions will be made for AD and Depression. And these, too, will be supported through modelling. Whether there is a relationship between AD and Depression will also be answered through analysis. Using the information obtained from the literature research conducted at the beginning of the project, at least five hypotheses regarding the factors for AD

and Depression will be proposed. The validity or invalidity of the hypotheses will be explained due to the tests performed.

In the second part of the project, a user interface will be designed to present the data mining study will be made at first part to the user. Differences from the current systems mentioned the section 2 also user interface opens the way for new analyses. The user interface will be able to load datasets to model by the user and analyse loaded datasets' relationship with AD and Depression. The system processes the loaded datasets and presents the analyses with modellings. Thus, the user will be able to define whether the datasets can be a factor for AD and Depression due to the study. The user can propose hypotheses in line with the datasets loaded. According to the result, when the system tests hypotheses, the user can conclude whether their hypotheses are valid or invalid. The user can make some predictions with the loaded datasets using the model. User can also save their studies into the model.

3.2. Functional Requirements

- The system should allow the users to load datasets into the model to identify factors of Alzheimer's Disease and Depression.
- The system should analyse the datasets that the user has uploaded and model the analyses; the user can determine the factors of AD and Depression using the results of the analyses.
- The system should test the user's hypotheses by using datasets loaded.
- The system should make predictions by using datasets loaded.
- The system allows the user to save their studies using the Data Mining Model for Factors of Alzheimer's and Depression.
- Users should be able to view AD and Depression rates based on country.
- Users should be able to view analyses and modellings of analyses obtained from datasets about factors of AD and Depression.
- Users should be able to observe the relationship between AD and Depression by analyses.
- Users should be able to view the predictions made for AD and Depression.
- Users should be able to observe the validity or invalidity of hypotheses proposed at the beginning of the project by the result of the tests.

3.3. Non-functional Requirements

3.3.1. Usability

A user will be able to view data mining studies about factors of AD and Depression and investigate other factors that may cause AD and Depression by loading datasets using analyses with modellings, conclude tested proposed hypotheses and make predictions easily. And also save their studies using the model easily so the model will be usable for the users.

3.3.2. Reliability

The system must be stable, so when the user does something wrong, the system works and consistently performs without failure.

3.3.3. Performance

The system will be fast enough that the user cannot wait long for saving their work. The system should also quickly present analyses, modelling, and tests to the user.

3.3.4. Supportability

Because this project is a research-based analysis and web-based project, there is no concern for supportability.

3.3.5. Implementation

The system will be implemented in Python as a programming language because this language is more suitable for this project.

3.3.6. Interface

There is a user interface where the user can view data mining studies about factors of AD and Depression and investigate other factors that may cause AD and Depression by loading datasets, making analyse with modelling with loaded datasets, concluding tested proposed hypotheses, and making predictions. And save their work using the model.

3.3.7. Packaging

Because this project is a research-based analysis and web-based project, there is no concern for packaging.

3.3.8. Legal

This project will use open-source datasets and open-source libraries.

3.4. System Models

3.4.1. Scenarios

Scenario Name: *ViewTheDataMiningStudy*

Participating Actor Instance: *Elif: User*

The Flow of Events:

1. Elif wants to do research where she can investigate the factors of AD and Depression. She opens The Data Mining Model for Factors of Alzheimer's Disease and Depression and examines the model.
2. The Data Mining Model for Factors of Alzheimer's Disease and Depression presents analysis studies made with datasets obtained from the most critical factors of AD and Depression, modellings of analyses expressed, which factors play a role in the determination of AD and Depression as a result of analysis studies, the relation between AD and Depression, predictions, proposed hypotheses about AD and Depression and the hypotheses about factors of AD and Depression, the tests of the hypotheses and results of the hypotheses.

Scenario Name: *ChangeParameters*

Participating Actor Instance: *Elif: User*

The Flow of Events:

3. Elif wants to change parameters, such as she does not want to see world data visualisation for a dataset only wants to see a specific country data visualisation for a dataset in the Data Mining Model for Factors of Alzheimer's Disease and Depression.
4. The Data Mining Model for Factors of Alzheimer's Disease and Depression offers the option of changing the parameter and presenting the data visualisation according to that parameter.

Scenario Name: *Load*

Participating Actor Instance: *Elif: User*

The Flow of Events:

1. Elif wants to make new analyses on the factors that can cause AD and Depression using The Data Mining Model for Factors of Alzheimer's Disease and Depression that she is examining. Elif loads the dataset she wants to analyse into the Data Mining Model for Factors of Alzheimer's Disease and Depression.
2. The Data Mining Model for Factors of Alzheimer's Disease and Depression process the dataset and presents the table-view of the dataset.

Scenario Name: *AnalyseWithMatrixAndHeamap*

Participating Actor Instance: *Elif: User*

The Flow of Events:

1. Elif wants to analyse the loaded dataset and the relationship between the loaded dataset and AD and Depression, so she wants to investigate the correlation between the loaded dataset and AD and Depression.
2. The Data Mining Model for Factors of Alzheimer's Disease and Depression present the correlation matrix and heatmap of loaded data and AD and Depression.

Scenario Name: *AnalyseWithModellings*

Participating Actor Instance: *Elif: User*

The Flow of Events:

1. Elif wants to analyse the loaded dataset and the relationship between the loaded dataset and AD and Depression with modelling.
2. The Data Mining Model for Factors of Alzheimer's Disease and Depression presents the analysis models as line charts, bullet graphs, pyramids graphs, etc.

Scenario Name: *TestHypotheses*

Participating Actor Instance: *Elif: User*

The Flow of Events:

1. Elif wants to propose some hypotheses and test her hypotheses according to the dataset loaded by using the Data Mining Model for Factors of Alzheimer's Disease and Depression so she can conclude her hypotheses.
2. The Data Mining Model for Factors of Alzheimer's Disease and Depression tests the hypotheses and presents the results of the hypotheses with modellings.

Scenario Name: *Predict*

Participating Actor Instance: *Elif: User*

The Flow of Events:

1. Elif wants to predict according to the dataset loaded by using the Data Mining Model for Factors of Alzheimer's Disease and Depression.
2. The Data Mining Model for Factors of Alzheimer's Disease and Depression predicts using regression and presents the result with modellings as scatter graphs, etc.

Scenario Name: *Save*

Participating Actor Instance: *Elif: User*

The Flow of Events:

1. Elif wants to save her studies.
2. The Data Mining Model for Factors of Alzheimer's Disease and Depression has a button to save the things that the user did.

3.4.2. Use case model

Use Case Name: ViewDataMiningStudy

Participating Actors: User

The Flow of Events:

1. The user opens the Data Mining Model for Factors of Alzheimer's Disease and Depression and investigates the model.
2. The system presents analysis studies made with datasets obtained from the most critical factors of AD and Depression.
3. The system presents analyses with modellings expressed with datasets obtained from factors of AD and Depression.
4. The system presents which factors play a role in determining AD and Depression due to analysis studies.
5. The system presents the relation between AD and Depression.
6. The system presents predictions.
7. The system presents proposed hypotheses about AD and Depression and the hypotheses about factors of AD and Depression, the tests of the hypotheses and the results of the hypotheses.

Entry Condition:

The user opens the model.

Exit Condition:

The System presents researched-based data mining studies in the model.

Quality Requirements:

The system presents data mining studies in a maximum of 3 seconds for each.

Use Case Name: ChangeParameters**Participating Actors:** User**The Flow of Events:**

1. The user change parameters, such as she does not want to see world data visualisation for a dataset only wants to see a specific country data visualisation for a dataset in the Data Mining Model for Factors of Alzheimer's Disease and Depression.
2. The system allows the user to change the parameter and present the data visualisation according to that parameter.

Entry Condition:

The user changes a parameter in the model.

Exit Condition:

The System visualises the dataset according to the changed parameter.

Quality Requirements:

The system visualises the changed parameter in a maximum of 2 seconds.

Use Case Name: Load**Participating Actors:** User**The Flow of Events:**

3. The user loads a dataset to analyse factors that can cause AD and Depression.
4. The system allows the user to load datasets in the model, process the dataset, and present the dataset's table view.

Entry Condition:

The user loads datasets into the model.

Exit Condition:

The System process and presents datasets correctly.

Quality Requirements:

The system process and presents datasets in a maximum of 2 seconds.

Use Case Name: AnalyseWithCorrelationAndHeatmap

Participating Actors: User

The Flow of Events:

1. The user analyses the loaded dataset and the relationship between the loaded dataset and AD and Depression.
2. The system presents the correlation matrix and heatmap of loaded data and AD and Depression.

Entry Condition:

The user makes analyses.

Exit Condition:

The system presents analyses with the correlation matrix and heatmap.

Quality Requirements:

The system presents a correlation matrix and heatmap in a maximum of 3 seconds.

Use Case Name: AnalyseWithModelings

Participating Actors: User

The Flow of Events:

3. The user analyses the loaded dataset and the relationship between the loaded dataset and AD and Depression with modelling.
4. The system presents the analysis models as line charts, bullet graphs, pyramids graphs, etc.

Entry Condition:

The user makes analyses with modelling.

Exit Condition:

The system presents analyses with modellings.

Quality Requirements:

The system presents analyses with modellings in a maximum of 3 seconds.

Use Case Name: TestHypotheses

Participating Actors: User

The Flow of Events:

1. The user proposes some hypotheses and tests them according to the dataset loaded, then the user can conclude about the hypotheses.
2. The system tests the hypotheses and presents the results of the hypotheses with modellings.

Entry Condition:

The user proposes hypotheses.

Exit Condition:

The system tests and presents the results of the tests.

Quality Requirements:

The system tests and presents the results of the tests in a maximum of 3 seconds.

Use Case Name: Predict**Participating Actors:** User**The Flow of Events:**

3. The user predicts according to the dataset loaded.
4. The system predicts using regression and presents the result with modellings as scatter graphs, etc.

Entry Condition:

The user makes a prediction.

Exit Condition:

The system predicts using regression.

Quality Requirements:

The system predicts using regression in a maximum of 3 seconds.

Use Case Name: Save**Participating Actors:** User**The Flow of Events:**

1. The user clicks the button to save the results of studies.
2. The system allows for saving users' results of studies.

Entry Condition:

The user clicks the button.

Exit Condition:

The System saves users' results.

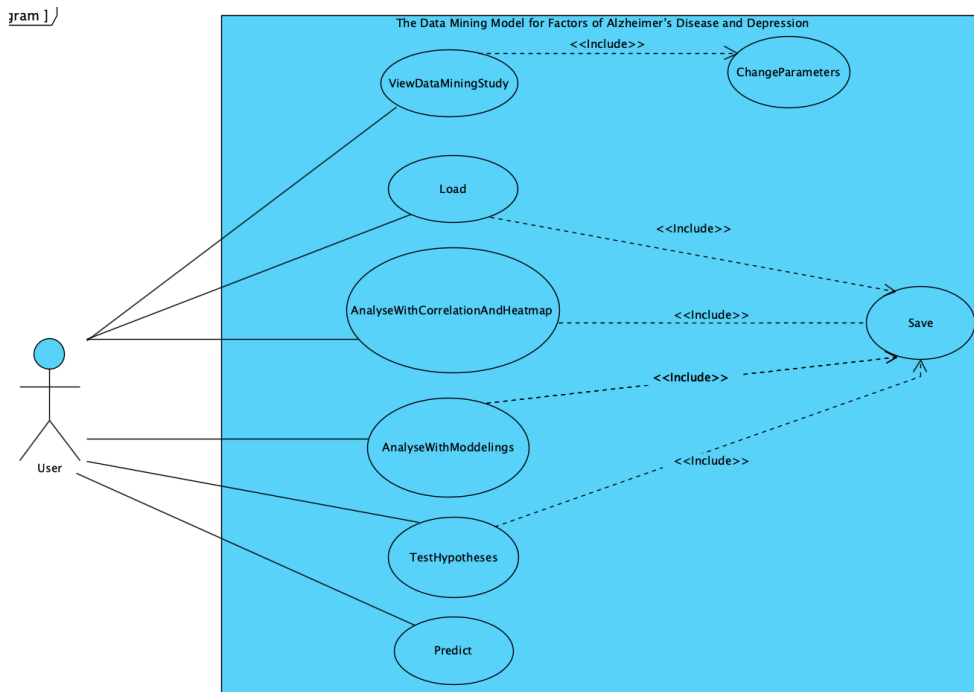
Quality Requirements:

The system saves the results of a user in a maximum of 3 seconds.

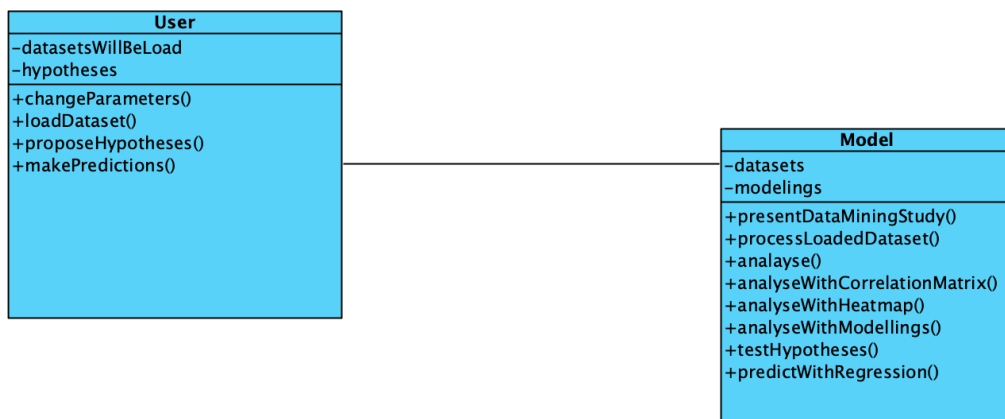
3.4.3. Object model

Models were also added to the ElifAkar.vpp project as a vpp file.

Model 1: Use Case Model



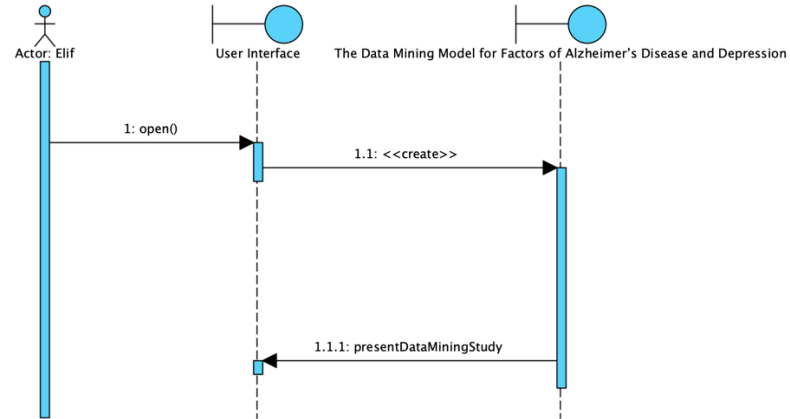
Model 2: Object Model



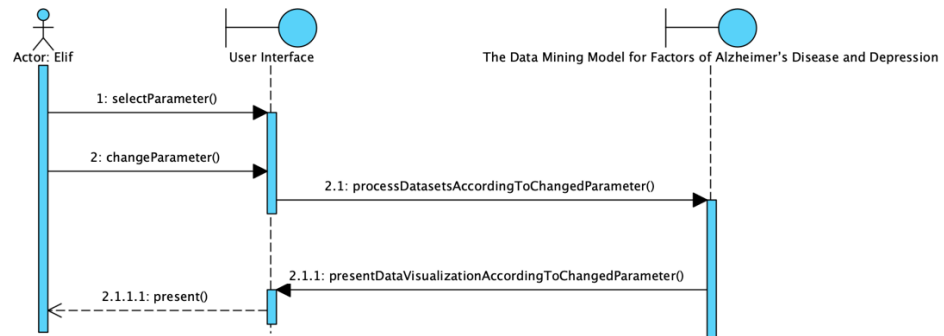
3.4.4. Dynamic model

Models were also added to the ElifAkar.vpp project as a vpp file.

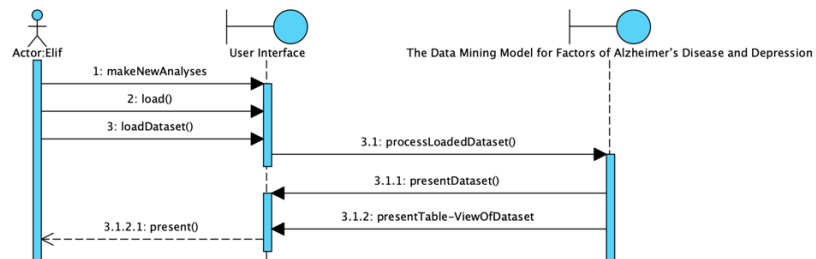
Model 1: ViewTheDataMiningStudy



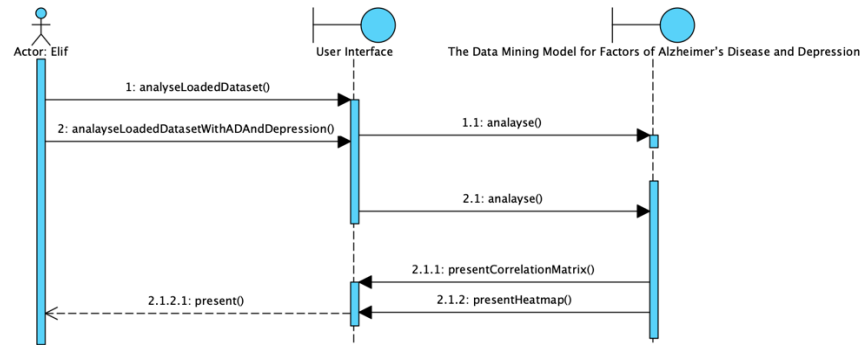
Model 2: ChangeParameters



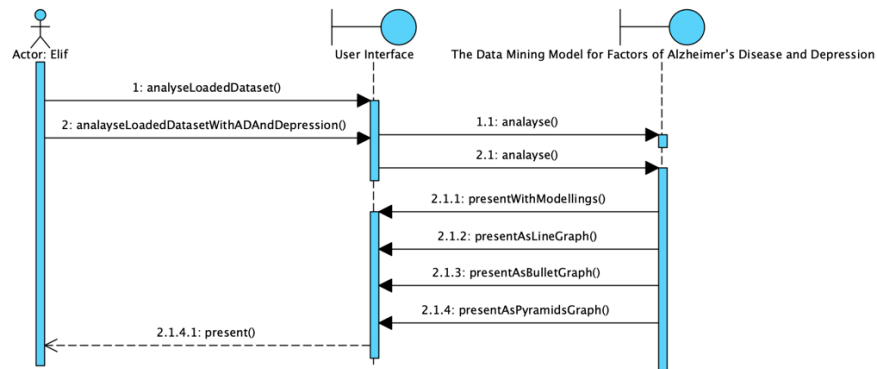
Model 3: Load



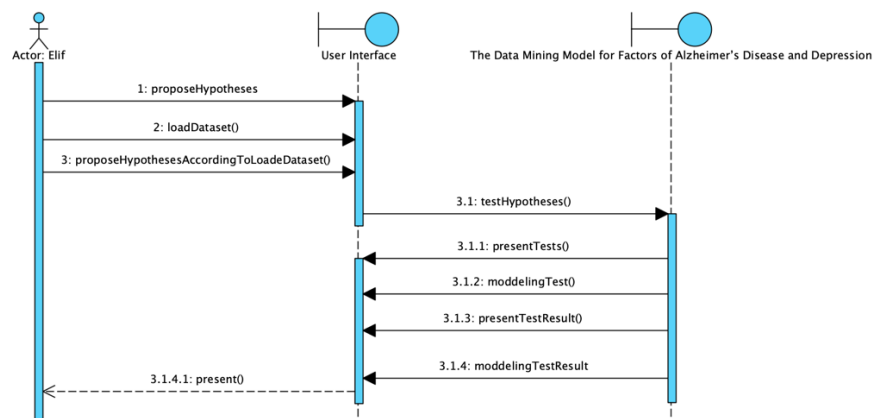
Model 4: AnalyseWithMatrixAndHeamap



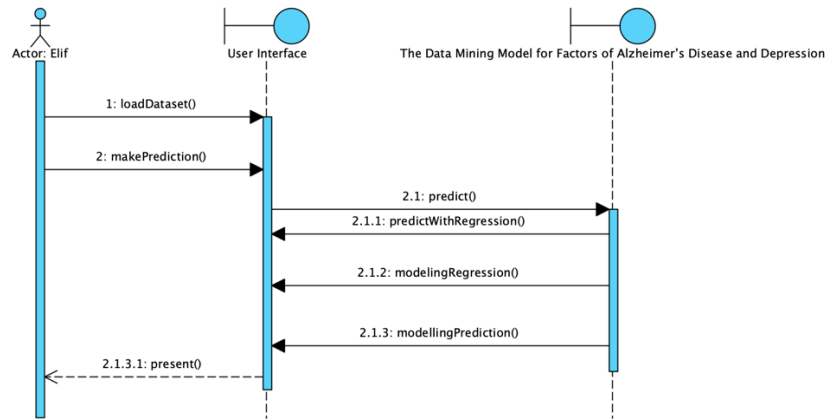
Model 5: AnalyseWithModellings



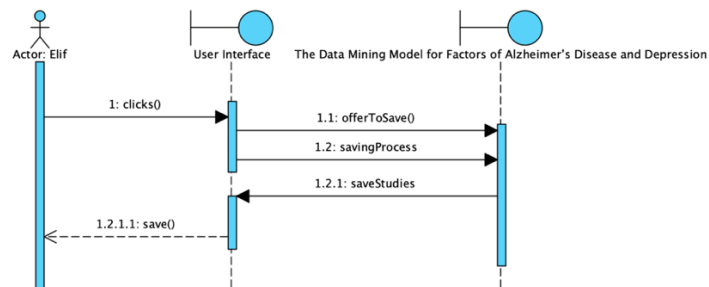
Model 6: TestHypotheses



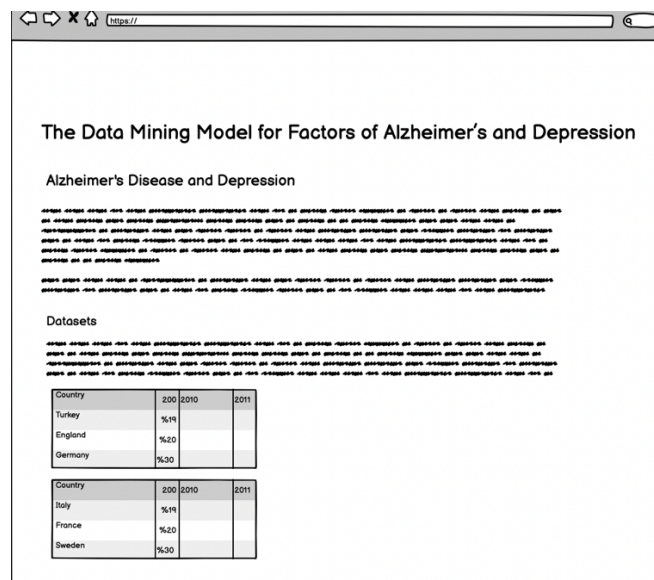
Model 7: Predict

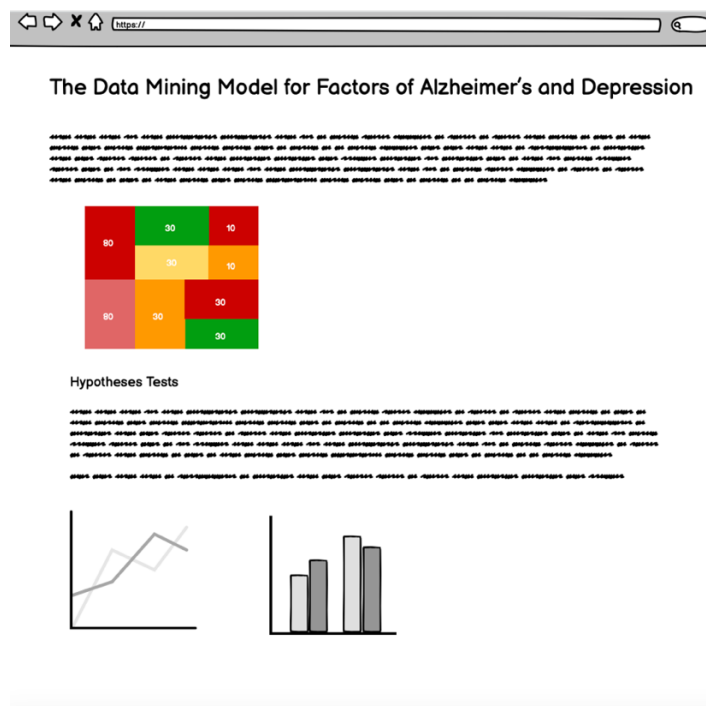
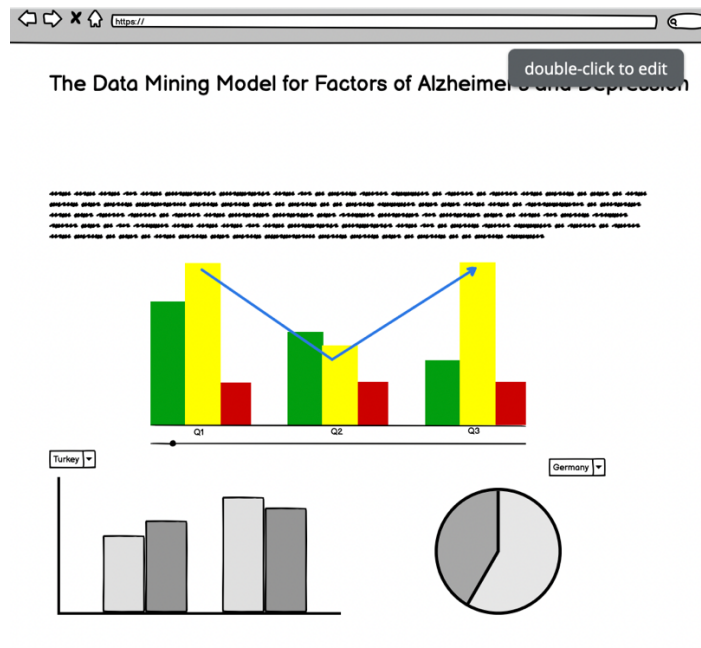


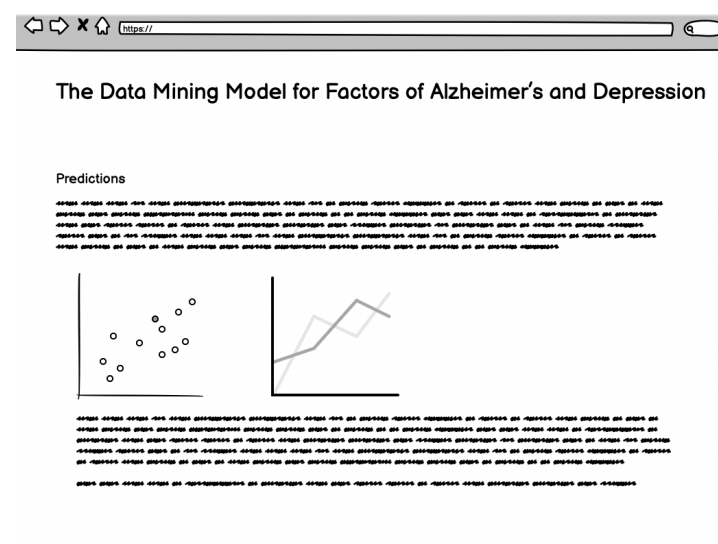
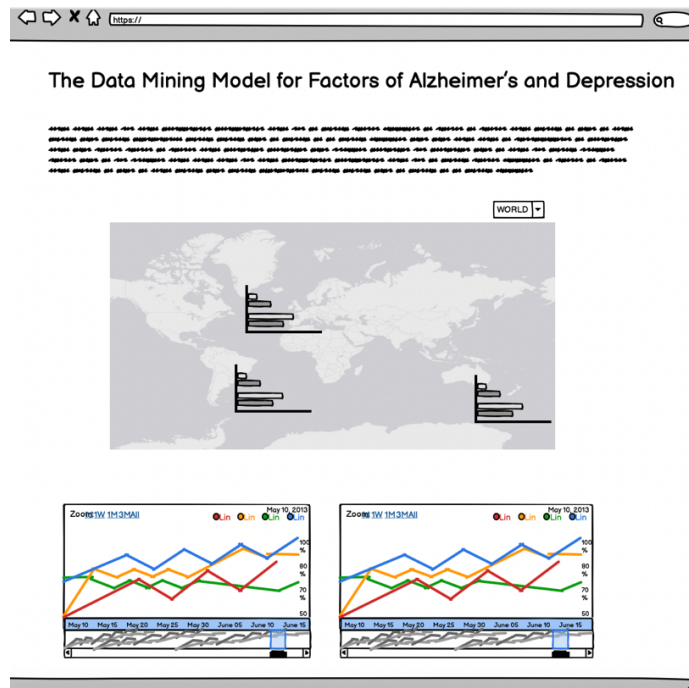
Model 8: Save



3.4.5. User interface—navigational paths and screen mock-ups







4. Glossary

AD: Alzheimer's Disease (AD) is neurodegeneration defined as a severe deterioration in the person's mental, physical, and behavioural functions due to disorders in the function of the brain. AD is one of the most common causes of dementia.

WHO: The World Health Organization (WHO) is the United Nations' specialised agency responsible for international public health.

5. References

1. WHO, Dementia: A Public Health Priority. World Health Organization and Alzheimer's Disease International (2012),
<https://www.who.int/publications/i/item/dementia-a-public-health-priority>
2. WHO, Depression and Other Common Mental Disorders: Global Health Estimates (2017),
<https://www.who.int/publications/i/item/depression-global-health-estimates>
3. Our World in Data, Mental Health (published 2018, updated 2021),
<https://ourworldindata.org/mental-health>
4. World Population Review, Depression Rates by Country (2022),
<https://worldpopulationreview.com/country-rankings/depression-rates-by-country>
5. Risk Factors and Identifiers for Alzheimer's Disease: A Data Mining Analysis (2014),
https://link.springer.com/chapter/10.1007/978-3-319-08976-8_1
6. WHO, WHO Coronavirus (COVID-19) Dashboard (2022),
<https://covid19.who.int>