SYSTEM DESIGN DOCUMENT

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

*by **Elif AKAR- 217CS2012***

<Submission Date: 24.04.2022>

Prepared for
COMP4902 Graduation Design Project

IŞIK UNIVERSITY
COMPUTER
SCIENCE AND
ENGINEERING

Table of Contents

SYSTEM DESIGN DOCUMENT

1. Introduction

Alzheimer's Disease (AD) is common neurodegeneration defined as a severe deterioration in a person's mental, physical, and behavioural functions due to disturbances in the brain. Depression is a common severe mood disorder characterised by moods such as sadness, guilt, low self-esteem, insomnia and loss of appetite, fatigue, and poor concentration that will affect our daily lives. In this context, a research-based data mining study was designed in The Data Mining Model for Factors of Alzheimer's and Depression Project, focusing on possible factors that may cause these two common mental illnesses. I will have created a model in which we can define AD and Depression with the analysis, hypothesis tests, and predictions to be made using datasets obtained from the factors decided from the study and research. In this way, users will be able to upload their datasets to this model, perform their analyses, and save their analyses, ultimately leading to new analyses defining the model AD and Depression.

1.1. Purpose of the System

The primary purpose of the system is to enable users to make their analysis studies for the identification of AD and Depression with the user interface created using The Data Mining Model for Factors of Alzheimer's and Depression and to enable them to save these studies and to contribute to new analysis studies thanks to this model. For this purpose, analysis studies, analysis models, the project and analysis studies are to be carried out with the data sets obtained based on countries from at least five critical factors that may cause AD and Depression, which were decided as a result of the research, in The Data Mining Model for Factors of Alzheimer's and Depression. The hypotheses produced together and the results obtained by testing these hypotheses will be presented to the user with the relationship between AD and Depression and the predictions made for these two mental illnesses.

1.2. Design Goals

The Data Mining Model for Factors of Alzheimer's and Depression is a research-based and web-based data mining project. To achieve this purpose, our design goals are:

1. Usability
The Data Mining Model for Factors of Alzheimer's and Depression should be a model developed for users can easily understand the analysis, hypothesis tests and results of the tests, and future predictions in the data mining work to be done in the first phase of the project. And in this context, a user interface should be created where users can easily make their analysis on the subject by using the presented data mining model.

2. Open-Source and Interactive
The Data Mining Model for Factors of Alzheimer's and Depression is an open-source project. Accessing the model starts uploading the project to GitHub, a cloud environment. All

users can access the model by downloading the model from this environment and opening the model on appropriate platforms. Also, The Data Mining Model for Factors of Alzheimer's and Depression is an interactive project. Users can make parameter changes on tables, graphics and maps on the data mining work made and presented to the user with the user interface and can reach the results they want to view. Users can run their analysis studies using the model and save these studies with the created user interface.

3. Research Focus

All users can perform analysis studies on AD and Depression based on the data mining work created and presented to the user using the user interface developed in The Data Mining Model for Factors of Alzheimer's and Depression project, and they can record these studies. Users can repeat their analysis and reach the same result in each iteration, *making* the project reliable.

4. Research -Base and Analyse

The Data Mining Model for Factors of Alzheimer's and Depression is a research-based project. Analysis studies, hypothesis tests and predictions made with datasets obtained from at least *the five* most *critical* factors that may cause Project AD and Depression are presented to users through the user interface. Users can upload datasets to the model and conduct their analysis to investigate factors that can cause AD and Depression.

5. Performance

The speed of the model's performance is essential in terms of quickly presenting the analysis, modelling, tests, and predictions that the user will make using the model to the user. In addition, the model will be so fast that the user does not have to wait long to save these works.

1.3.   Definitions, Acronyms, and Abbreviations

AD – Alzheimer's Disease

1.4.   References

1. The Data Mining Model for Factors of Alzheimer's and Depression Requirement Analysis Document

2. Literature Research

Factors that can cause AD and Depression and the relationship between AD and Depression have been the subject of many research projects before. As a result of research, the factors that can cause AD and Depression and the relationship between AD and Depression have been presented in various articles. For this reason, I benefited from these articles that I reached by doing a literature search at the beginning of the project to determine at least the five most critical factors that we will use as a basis for deciding on the datasets that form the basis of The Data Mining Model for Factors of Alzheimer's and Depression project.

The aim of the Risk Factors and Identifiers for Alzheimer's Disease: A Data Mining Analysis study, published in 2014 by Gürdal Ertek, Bengi Tokdil, and İbrahim Günaydın, is to analyse

the risk factors of AD and to determine the tests that can help the diagnosis of AD based on these risk factors [1]. In the Data Mining Model, data were obtained from the Open Access Series of Imaging Studies (Marcus et al., 2010). In the dataset, whether the patient had AD or not was stated as dementia and non-dementia [1]. Attributes (factors) in the data set analysed in the study; Group, MRIID, SubjectID, Visit, MRDelay, CDR, Gender, Age, Education level, SES, MMSE, eTIV, nWBV, ASF are presented in such a way as to provide the relevant value ranges [1]. Analysis studies were carried out using Orange and Tableau data mining software, examining attributes both among each other and among patients with or without dementia [1]. Some conclusions were reached at the end of the analysis; if women have an MMSE greater than 28, there is an 84.6% probability that they do not have dementia; less educated subjects in males show early signs of dementia; For men with EDUC >15, MMSE >28, and ASF >0.928, the probability of not having dementia is 88.9% if age is less than or equal to 76; dementia was observed in half of the female trials; for men, nWBV>0.680 indicates a greater probability of risk; Alzheimer's risk is higher in people with a college degree or higher; It was concluded that university graduate women are at higher risk of developing Alzheimer's disease than university graduate men [1].

Published by the American Journal of Epidemiology in 2000, The aim of the article Education and the Risk for Alzheimer's Disease: Sex Makes a Difference, EURODEM Pooled Analyses is to examine the relationship between years of schooling and dementia and AD [2]. The data in the study were obtained from European population-based follow-up studies [2]. In the analysis study, education level was categorised as low, medium, or high according to years of education [2]. The study estimated age, sex, work centre, smoking status, and self-reported myocardial infarction and stroke using Poisson's regression for relative risks (95% confidence intervals) alongside education level [2]. In the statistical analyses performed in the study, it was observed that the relative risk for dementia and Alzheimer's disease was marginally increased for those in the low-education group and those in the middle-education group when compared to the high-education group [2]. In addition, it was understood that there was a significant interaction between gender and education level in terms of dementia and AD risk [2]. Some results we will reach from the analysis; there is an increased risk for dementia, particularly AD, for women, but not for men, associated with a reduced number of school years; The risk of AD is higher in women than in men; however, the fact that in the data collected women are, on average, less educated and therefore at higher risk for AD, may be misleading [2].

The purpose of the article Gender Differences in Causes of Depression, written by Marta Elliott PhD, in 2001, is to analyse gender differences in the causes of depression [3]. In the study, it is adopted to look at the stress process perspective of the individual to analyse the gender differences in the causes of depression [3]. The study took stress factors and sources as mediators of the stress/depression relationship [3]. In this context, the relationship between depression/gender has been tested by hypotheses that women are more exposed to stress factors and more vulnerable than men; women benefit more from socially supportive relationships and suffer more from conflictual relationships than men [3]. Survey data from 45-74-year-old Nevada residents collected in 1997 were used in the analysis, and ordinary least squares regression was used to test the stress process model [3]. Based on the relationships between stress/depression in the model used, hypotheses have been proposed to explain the relationship between depression and gender [3]. The study argues that SES reveals and makes people

vulnerable to stress factors in different ways. Testing and predicting depression with a stress process model is covered. Some of the results obtained from the study; are that women have low SES. Therefore, they are exposed to more stress, which is a critical source of their tendency to depression; This indicates that we can infer gender inequality in socioeconomic status. Women are more likely to suffer from economic hardship, which causes stress and depression [3].

The article titled Risk Factors for Depression Among Elderly Community Subjects: A Systematic Review and Meta-Analysis, published by Martin G. Cole and Nandini Dendukuri in 2003 aims to determine what side factors may influence the progression of major depression with increasing age [4]. The study compared adults with depressive symptoms, with or without the depressive disorder, with people with chronic medical conditions such as heart and lung disease, hypertension, diabetes, and arthritis [4]. In addition, topics such as depression and benefiting from medical services and health care are also covered [4]. Age, sleep disorder, and gender are also among the subjects investigated [4]. Data were summarised from several reports with information on the age of issues, the proportion of males, criteria for depression, initial exclusion criteria, length of follow-up, number of cases of depression, and risk factors [4]. As a result of the analysis, there are five risk factors for depression in the elderly in the community, including age, sleep disturbance, disability, previous cases of depression, and female gender [4].

The article titled Food Combination and Alzheimer's Disease Risk: A Protective Diet, published in 2010 by Yian Gu, Jeri W. Nieves, Yaakov Stern, Jose A. Luchsinger, and Nikolaos Scarmeas aims to make sense of the relationship between the variety of nutrients associated with AD [5]. The data were obtained through surveys and analysis by the Channing Laboratory, Cambridge, Massachusetts [5]. As a result of the comments made in the study; It was concluded that vitamin E could prevent AD with its substantial antioxidant effect; showing higher consumption of certain foods (salad dressing, nuts, fish, tomatoes, poultry, cruciferous vegetables, fruits, dark and leafy greens) and less consumption of others (high-fat dairy, red meat); (organ meat and butter) may be associated with a reduced risk of developing AD through a more favourable nutrient profile (vitamin E and folate intake) [5].

The article titled Risk factors for depression in elderly people: a prospective study, published in 1992 by Green BH, Copeland JRM, Dewey ME, Sharma V, Saunders PA, Davidson IA, Sullivan C, McWilliam C. aims to examine the risk factors and to conclude the relationship between them [6]. The risk factors mentioned in the study include age, gender, marital status, socioeconomic status, physical illness, and disability [6]. One of the hypotheses emphasises the high rate of depression in women-focused on reproductive years and stresses the role of marital status in this gender difference [6]. Data are from a health study conducted in Liverpool [6]. Some of the results obtained in the survey; depression is predicted in smokers, but not necessarily a history of smoking; relations with friends and family, having a psychiatric history in the family, being over 65 years old do not have an essential role in the development of depressive illness; about 40% of depressed cases were found to have some form of cardiovascular disease; Family history of depression was found in 7 out of 44 depressive patients, but there was no significant difference when compared with the control group; log-linear modelling and various models have been used and tested to determine the independence

of risk factors and whether there is any interaction; Lack of life satisfaction, smoking and loneliness have been confirmed as significant risk factors [6].

2.1. Data Mining Study

Based on the articles obtained from the literature research, the data mining study, the first part of The Data Mining Model for Factors of Alzheimer's and Depression project, will be started by determining at least the five most important factors that can cause AD and Depression. By determining the factors, at least five hypotheses will be put forward by examining the factors between AD and Depression, between AD and Depression or by examining the factors within themselves. After the factors are determined and hypotheses are put forward, datasets will be obtained based on factors based on countries. The analysis part of the data mining work will begin with finding the datasets and making them ready for analysis. The analysis will start with the fundamental analysis of the datasets, continue with the analysis of the correlation matrix and heatmap, and finally, the analysis will be done with the models; datasets and the relationships between AD and Depression and the relationship between AD and Depression and the relationships within the factors themselves will be concluded. After the analysis study, the hypotheses put forward will be tested, and the validity or invalidity of the hypotheses will be concluded. Finally, predictions will be made for AD and Depression. Thus, the first part of The Data Mining Model for Factors of Alzheimer's and Depression project will end. The data mining work will be developed in the project, as mentioned, based on the articles obtained from the literature research.

2.2. Definitions, Acronyms, and Abbreviations

Group – Demented, non-demented, or converted
MRIID – The test ID
Visit – Visit the subject
MRDelay – The delay of subject
CDR – Clinical Dementia Rating
EDUC – Education level
SES – Socioeconomic status
MMSE – Mini-Mental State Examination
eTIV – Estimated total intracranial volume
nWBV – Normalized whole-brain volume
ASF – Atlas Scale Factor

3. Proposed Software Architecture

3.1. Overview

The Data Mining Model for Factors of Alzheimer's and Depression is a research-based and web-based project. The model will be developed with a data mining study, and the user interface will be created and presented to the user. Therefore, the architecture of the project grows in two stages.

The first phase is where the model is created with a data-mining study. How to develop this phase was mentioned in section 2.1.

The second stage is the part of the project developed differently from the articles obtained in the literature research. The purpose of this phase covers how the data mining work based on the model presented to the user and how the user benefits from it. For this reason, a user interface will be developed. With the interface to be created, the user can access the model. By making parameter changes to the visualisations of the data-mining operation, the user can switch between analysis graphs, maps or tables and reach the desired result. The user can make their analyses using the model created in the data mining study and can define AD and Depression with the analyses they will make. In this work order, the user starts to load the data set into the model. The study follows the stages of the data mining work done in the first phase of the User model sequentially. The user can test the hypotheses put forward in line with the data set with the test methods made in the first stage and reach the result. The user can create new predictions using the prediction section made in the first stage. And it can save these studies using the model.
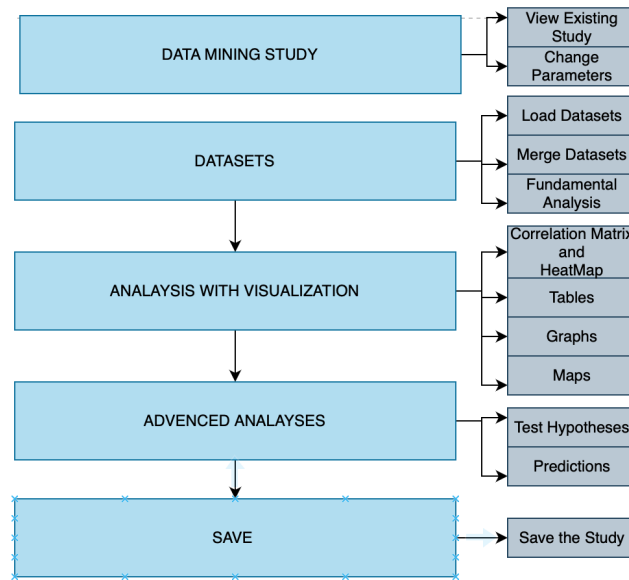

Figure 1: *User Interface Elements Model*

3.2.   System Decomposition

The Data Mining Model for Factors of Alzheimer's and Depression consists of five subsystems:

- **Data Mining Study Subsystem**

  The Data Mining Study subsystem includes the ability to view the data mining study called The Data Mining Model for Factors of Alzheimer's and Depression by accessing it via the user interface. Users can make parameter changes to the visualisations in their work and view the analysis model they want to obtain.

- **Datasets Subsystem**

  The Datasets subsystem contains a structure where they can upload the dataset for their analysis studies using The Data Mining Model for Factors of Alzheimer's and

Depression. It is also a subsystem that merges datasets and performs fundamental analysis on datasets.

- **Analysis with Visualization Subsystem**

  Analysis with Visualization subsystem includes a structure where users can analyse the datasets. They have loaded using The Data Mining Model for Factors of Alzheimer's and Depression by using correlation matrix and heatmap, creating tables, graphs, and maps, and obtaining analysis results in these ways.

- **Advanced Analysis Subsystem**

  The Advanced Analysis subsystem includes a structure where users will be able to test various hypotheses; they propose using datasets they have uploaded using The Data Mining Model for Factors of Alzheimer's and Depression; It contains a structure in which they can decide the validity or invalidity of the results of hypothesis tests and make predictions about AD and Depression.

- **Save Subsystem**

  The Save subsystem includes a structure for users to save their work using The Data Mining Model for Factors of Alzheimer's and Depression.
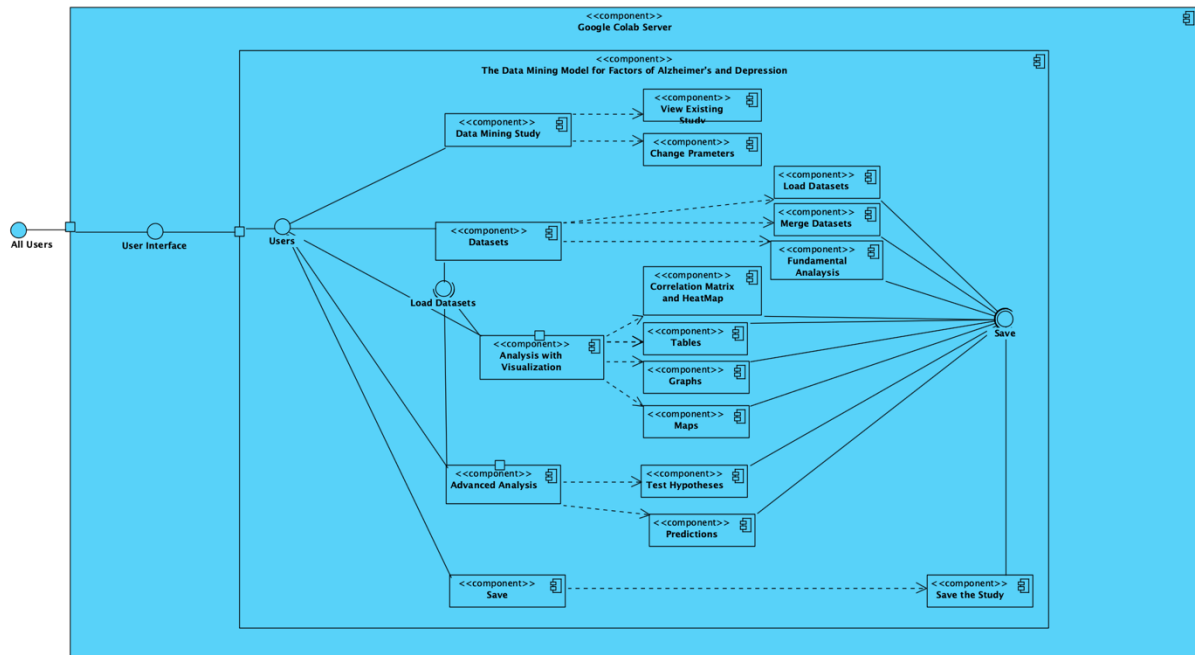

Figure 2: *Component Diagram Model*
(The model was also added to the ElifAkar_SDD.vpp project as a vpp file.)

3.3. Hardware/Software Mapping

The Data Mining Model for Factors of Alzheimer's and Depression is a cloud-based project, and the project is run on Google Colab. And the user interface is implemented on Colab.
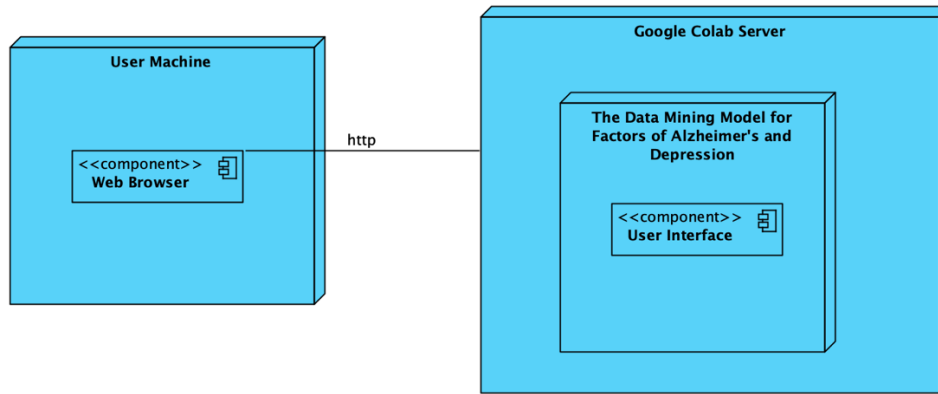
Figure 3: *Hardware/Software Mapping Model*
(The model was also added to the ElifAkar_SDD.vpp project as a vpp file.)

3.4. Persistent Data Management

The Data Mining Model for Factors of Alzheimer's and Depression is a cloud-based project. Users access the project via GitHub. All users can access the project, and no user password is required. Users do not need to be registered to view the model and work on the model; the system does not have a user registration feature. Users record their work on the model, and this record is not kept in a database. For these reasons, the project does not include a database.

3.5. Access Control and Security

The Data Mining Model for Factors of Alzheimer's and Depression project will be available to users via a GitHub cloud environment. Users will be able to access and work on the model by downloading the model from GitHub and opening it on the appropriate platform.

3.6. Global Software Control

All users will be able to access The Data Mining Model for Factors of Alzheimer's and Depression project through GitHub, a cloud environment. For the users to work on the project in an up-to-date manner and not to lose the correctness of the project, the data sets will be updated regularly, the updated data sets will be added to the model so that the model remains up-to-date, and the updated data sets will be presented to the user within the project.

3.7. Boundary Conditions

Start-up:

- Google Colab server start-up.
- The Data Mining Model for Factors of Alzheimer's and Depression is downloaded from the cloud environment of GitHub.
- Users examine the data mining work using the user interface and review the analysis studies by changing the parameters.
- Users load datasets into the model using the user interface.

10

– Users make and save their analysis work using the user interface.

Shutdown:

- Users shut down The Data Mining Model for Factors of Alzheimer's and Depression.
- Users' work using The Data Mining Model for Factors of Alzheimer's and Depression is saved.
- Google Colab server shut down.

Error Condition:

- If users do not load the data set in the model in the format given in the model, the model will provide an error.
- If the user opens the model on a platform other than Google Colab, the model may fail.

4. Subsystem Services

- **Data Mining Study Subsystem**
- *View Existing Study Service:* This service includes a structure where users can view the data mining work that forms the basis of The Data Mining Model for Factors of Alzheimer's and Depression, thanks to the user interface created.
-
- *Change Parameters Service:* This service provides a structure that allows users to view the analysis result they want to obtain by making parameter changes on the visualised analyses they consider in the data mining work that forms the basis of The Data Mining Model for Factors of Alzheimer's and Depression.

- **Datasets Subsystem**
- *Load Dataset Service:* This service provides where users can load their datasets to conduct analysis using The Data Mining Model for Factors of Alzheimer's and Depression.

- *Merge Dataset Service:* This service provides a structure for users to merge their loaded datasets using The Data Mining Model for Factors of Alzheimer's and Depression.

- *Fundamental Analysis Service:* This service provides a structure for users to perform fundamental analyses with the data sets they have loaded using The Data Mining Model for Factors of Alzheimer's and Depression.

- **Analysis with Visualization Subsystem**
- *Correlation Matrix and HeatMap Service:* This service provides a structure for users to analyse their loaded datasets with the correlation matrix and heatmap using The Data Mining Model for Factors of Alzheimer's and Depression.

11

- *Tables Service:* This service provides a structure for users to analyse the datasets they have uploaded using The Data Mining Model for Factors of Alzheimer's and Depression by creating tables.

- *Graphs Service:* This service provides a structure for users to analyse the datasets they load using The Data Mining Model for Factors of Alzheimer's and Depression by creating graphs such as line charts, bullet graphs, pyramids graphs, etc.

- *Maps Service:* This service provides a structure for users to analyse the datasets they have loaded using The Data Mining Model for Factors of Alzheimer's and Depression by creating maps.

- **Advanced Analysis Subsystem**
- *Test Hypotheses Service:* This service provides a structure where users can test and finalise their hypotheses based on the datasets they have loaded using The Data Mining Model for Factors of Alzheimer's and Depression.

- *Predictions Service:* This service provides a structure where users can make some predictions about the datasets they have loaded using The Data Mining Model for Factors of Alzheimer's and Depression.

- **Save Subsystem**
- *Save the Study Service:* This service provides a structure where users can save their work using The Data Mining Model for Factors of Alzheimer's and Depression.

5. References

1. 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
2. Education and the Risk for Alzheimer's Disease: Sex Makes a Difference. EURODEM Pooled Analyses (2000), https://academic.oup.com/aje/article/151/11/1064/87271
3. Gender Differences in Causes of Depression (2001), https://www.tandfonline.com/doi/abs/10.1300/J013v33n03_11
4. Risk Factors for Depression Among Elderly Community Subjects: A Systematic Review and Meta-Analysis (2003), https://ajp.psychiatryonline.org/doi/abs/10.1176/appi.ajp.160.6.1147
5. Food Combination and Alzheimer's Disease Risk: A Protective Diet (2010), https://pubmed.ncbi.nlm.nih.gov/20385883/
6. Risk factors for depression in elderly people: a prospective study (1992), https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1600-0447.1992.tb03254.x