IŞIK UNIVERSITY
COMPUTER
SCIENCE AND
ENGINEERING

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

Bachelor’s Thesis

Elif Akar
217CS2012

Supervised by
F. Boray Tek

June 2022

ABSTRACT

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.

ACKNOWLEDGEMENTS

I would like to present my special thanks to my supervisor **Assistant Professor Faik Boray Tek** for his guidance, help, guidance and patience in this process.

TABLE OF CONTENTS

LIST OF FIGURES

DEFINITIONS, ACRONYMS AND ABBREVIATIONS

- AD – Alzheimer’s Disease
- WHO – World Health Organization
- 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

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, The Data Mining Model for Factors of Alzheimer's and Depression Project focuses 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 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 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 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.

1.3. Objectives and Success Criteria of System

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.

2. Literature Review

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 [3]. 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 [3]. 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 [3]. 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 [3]. 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 [3].

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 [4]. The data in the study were obtained from European population-based follow-up studies [4]. In the analysis study, education level was categorised as low, medium, or high according to years of education [4]. 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 [4]. 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 [4]. In addition, it was understood that there was a significant interaction between gender and education level in terms of dementia and AD risk [4]. 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 [4].

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 [5]. 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 [5]. The study took stress factors and sources as mediators of the stress/depression relationship [5]. 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 [5]. 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 [5]. Based on the relationships between stress/depression in the model used, hypotheses have been proposed to explain the relationship between depression and gender [5]. 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 [5].

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 [6]. 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 [6]. In addition, topics such as depression and benefiting from medical services and health care are also covered [6]. Age, sleep disorder, and gender are also among the subjects investigated [6]. 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 [6]. 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 [6].

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 [7]. The data were obtained through surveys and analysis by the Channing Laboratory, Cambridge, Massachusetts [7]. 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) [7].

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 [8]. The risk factors mentioned in the study include age, gender, marital status, socioeconomic status, physical illness, and disability [8]. 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 [8]. Data are from a health study conducted in Liverpool [8]. 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 [8].

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.

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. Nonfunctional 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: *AnalayseWithModellings*
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 Models

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: AnalyseWithModdelings
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 Models

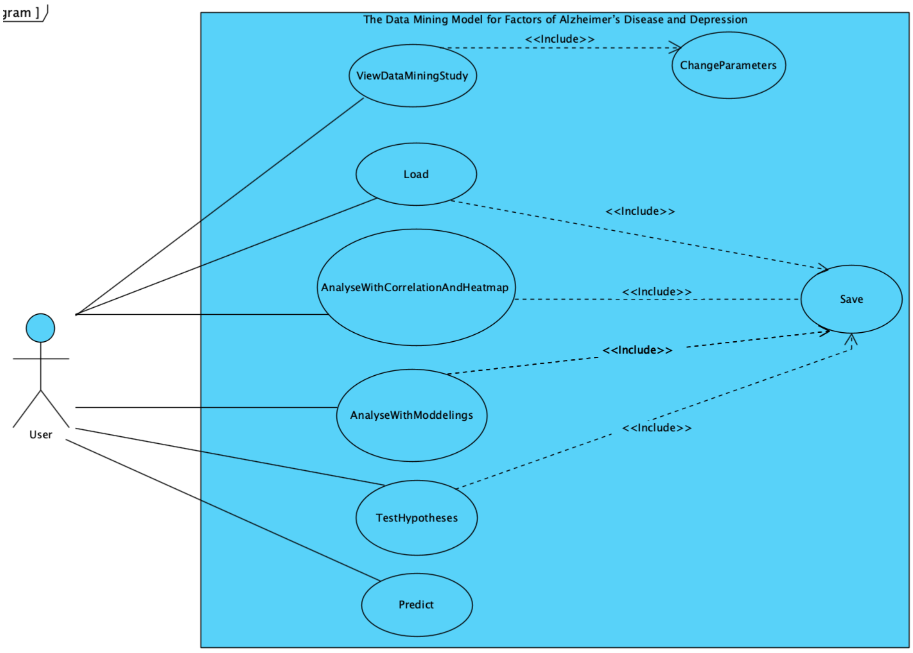


Figure 1 *Use Case Model*

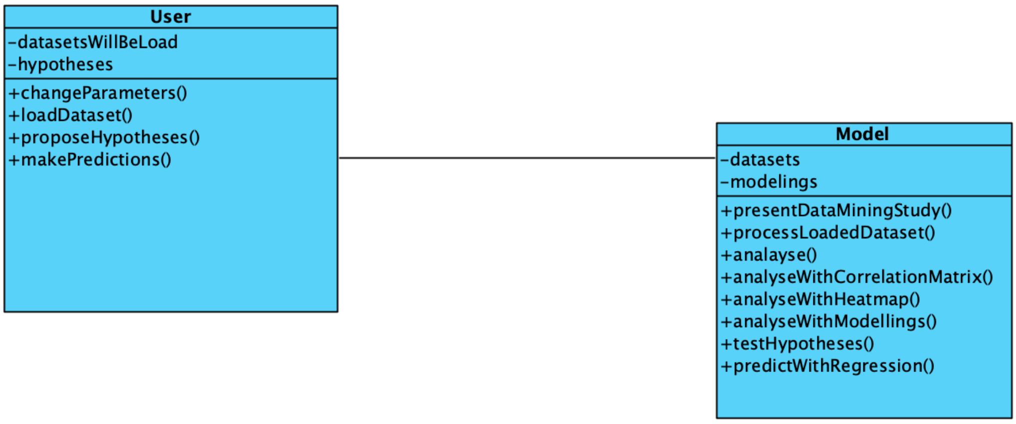
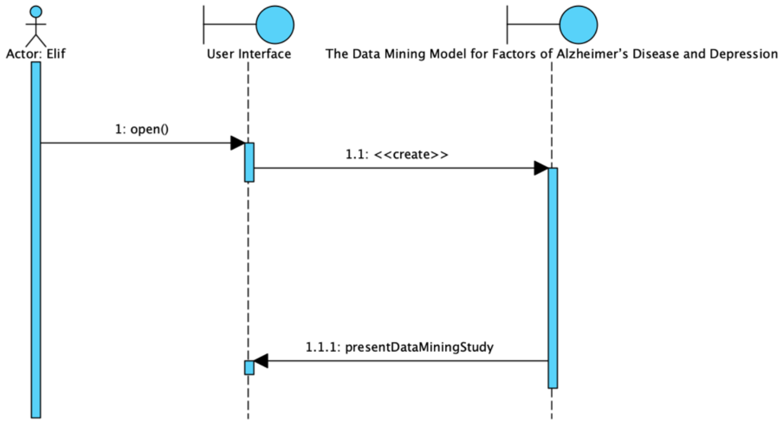


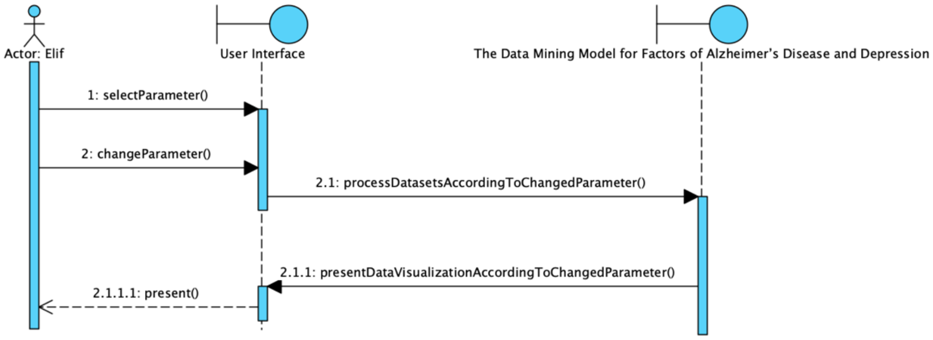
Figure 2 *Object Model*

3.4.4. Dynamic Models



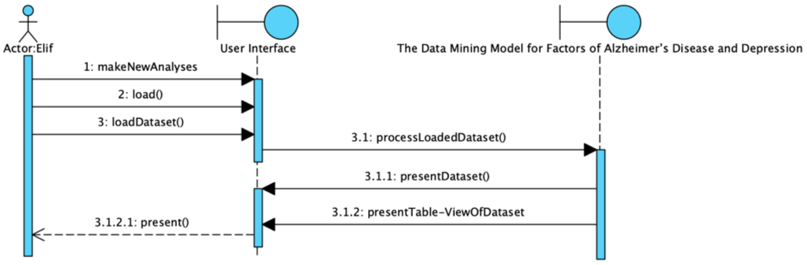
Actor: Elif
User Interface
The Data Mining Model for Factors of Alzheimer's Disease and Depression
1: open()
1.1: <<create>>
1.1.1: presentDataMiningStudy

Figure 3 *Dynamic Model 1: ViewTheD ataMiningStudy*



Actor: Elif
User Interface
The Data Mining Model for Factors of Alzheimer's Disease and Depression
1: selectParameter()
2: changeParameter()
2.1: processDatasetsAccordingToChangedParameter()
2.1.1: presentDataVisualizationAccordingToChangedParameter()
2.1.1.1: present()

Figure 4 *Dynamic 2: ChangeParameters*



Actor:Elif
User Interface
The Data Mining Model for Factors of Alzheimer's Disease and Depression
1: makeNewAnalyses
2: load()
3: loadDataset()
3.1: processLoadedDataset()
3.1.1: presentDataset()
3.1.2: presentTable-ViewOfDataset
3.1.2.1: present()

Figure 5 *Dynamic Model 3: Load*

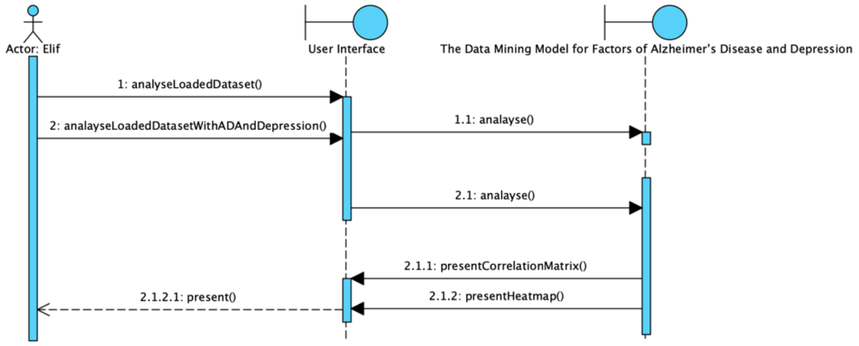

Actor: Elif
User Interface
The Data Mining Model for Factors of Alzheimer's Disease and Depression
1: analyseLoadedDataset()
2: analayseLoadedDatasetWithADAndDepression()
1.1: analayse()
2.1: analayse()
2.1.1: presentCorrelationMatrix()
2.1.2: presentHeatmap()
2.1.2.1: present()

Figure 6 *Dynamic Model 4: AnalyseWithMatrixAndHeamap*

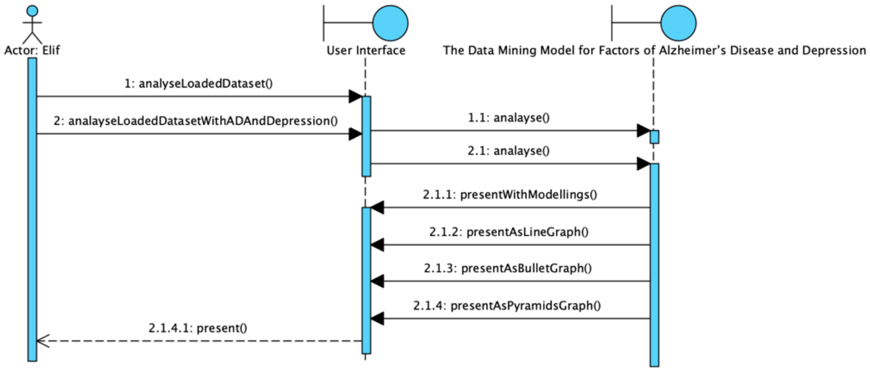

Actor: Elif
User Interface
The Data Mining Model for Factors of Alzheimer's Disease and Depression
1: analyseLoadedDataset()
2: analayseLoadedDatasetWithADAndDepression()
1.1: analayse()
2.1: analayse()
2.1.1: presentWithModellings()
2.1.2: presentAsLineGraph()
2.1.3: presentAsBulletGraph()
2.1.4: presentAsPyramidsGraph()
2.1.4.1: present()

Figure 7 *Dynamic Model 5: AnalyseWithModellings*

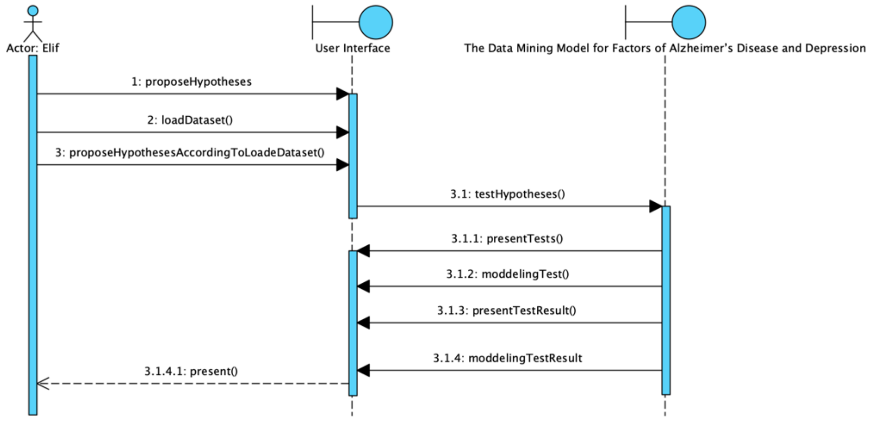

Actor: Elif
User Interface
The Data Mining Model for Factors of Alzheimer's Disease and Depression
1: proposeHypotheses
2: loadDataset()
3: proposeHypothesesAccordingToLoadeDataset()
3.1: testHypotheses()
3.1.1: presentTests()
3.1.2: moddelingTest()
3.1.3: presentTestResult()
3.1.4: moddelingTestResult
3.1.4.1: present()

Figure 8 *Dynamic Model 6: TestHypotheses*

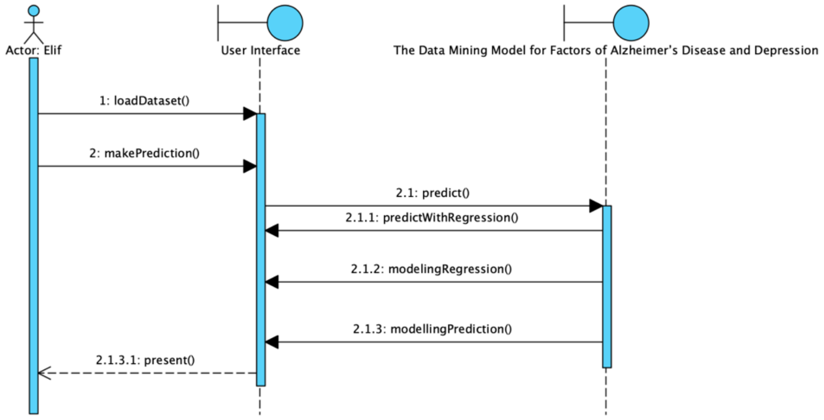

Actor: Elif
User Interface
The Data Mining Model for Factors of Alzheimer's Disease and Depression
1: loadDataset()
2: makePrediction()
2.1: predict()
2.1.1: predictWithRegression()
2.1.2: modelingRegression()
2.1.3: modellingPrediction()
2.1.3.1: present()

Figure 9 *Dynamic Model 7: Predict*

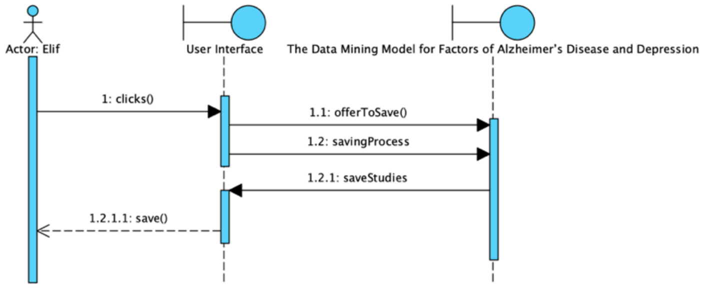


Figure 10 *Dynamic Model 8: Save*

3.5. Proposed System Architecture

3.5.1. 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.

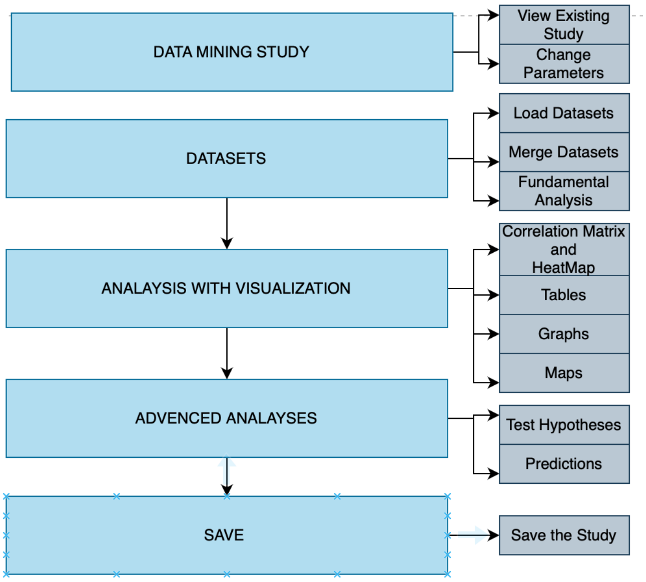


Figure 11 *User Interface Elements Model*

3.5.2. 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 12 Component Diagram Model

3.5.3. 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.

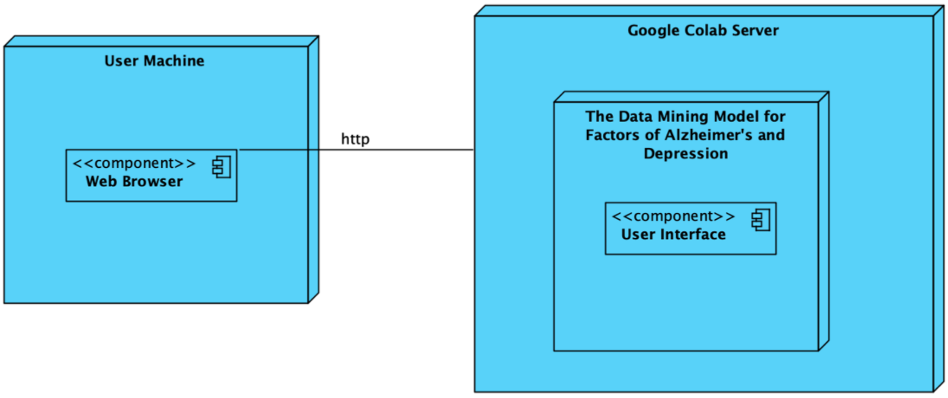


Figure 12 *Hardware/Software Mapping Model*

3.5.4. President 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.

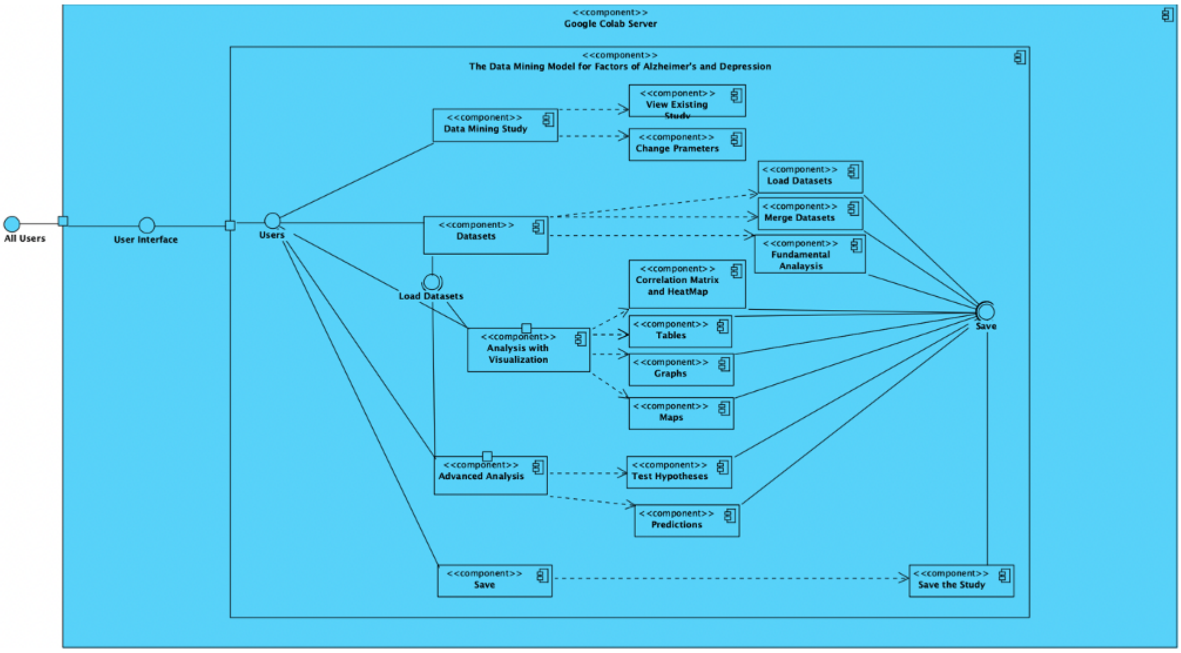


Figure 13 *Component Diagram Model*

3.5.5. 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.

3.5.6. 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.

3.5.7. 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.
– 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.

3.6. 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.

- *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.

4. Implementation, Hypotheses, Tests, Experiments

The software language used in The Data Mining Model for Factors of Alzheimer's and Depression Project is Python and it uses the libraries provided by Python. Since the project is a research-based data mining project, the implementation of the project consisted of two stages. In the first stage, data sets were obtained based on the factors determined as a result of the literature research. In the second stage, analyzes were carried out with the obtained data sets, visualizations were made and they were integrated with the user interface in order to be interactive with the user. The project was created using Google Colab.

4.1. Datasets

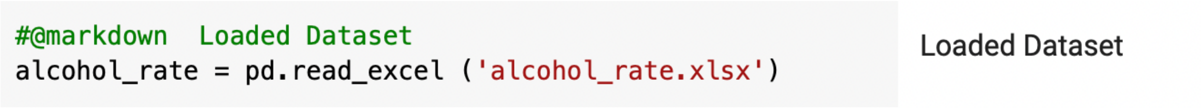
All of the data sets used in The Data Mining Model for Factors of Alzheimer's and Depression Project were taken from the same database as a result of my research in order to be consistent, and they were drawn and edited from the source used by the database.

- AD Dataset [9]
- Depression Dataset [10]

- Calorie Nutrition Dataset [11]
- Smoking Dataset [12]
- Depression by Age Dataset [13]
- Depression by Genre [14]
- GDP Dataset [15]
- Average Education Dataset [16]
- Alcohol Dataset [17]

4.2. Analyses Process, and User Interface

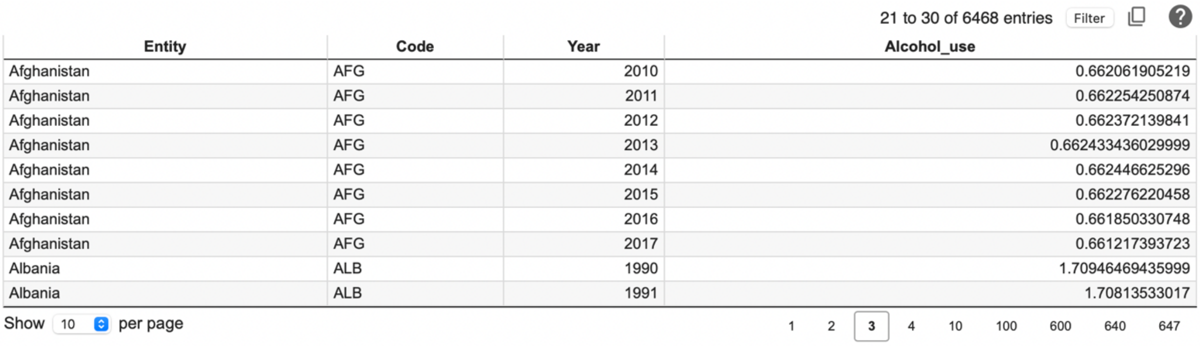
In The Data Mining Model for Factors of Alzheimer's and Depression Project, I started the data analysis process by analyzing the datasets that I determined as factors that can cause AD and Depression. I have read the files of the datasets, and these are the sections where users will upload their own datasets to the model. Before presenting the tables of the datasets to the user, I defined the datasets using the describe() and info() functions and presented the data tables to the user in an interactive way. Then, I presented the charts showing the values of the countries of that data set to the user with a user interface where the users can use the charts interactively. I did these operations for all factor datasets.



```
#@markdown  Loaded Dataset
alcohol_rate = pd.read_excel ('alcohol_rate.xlsx')
```

Loaded Dataset

Figure 14 *Load Dataset*

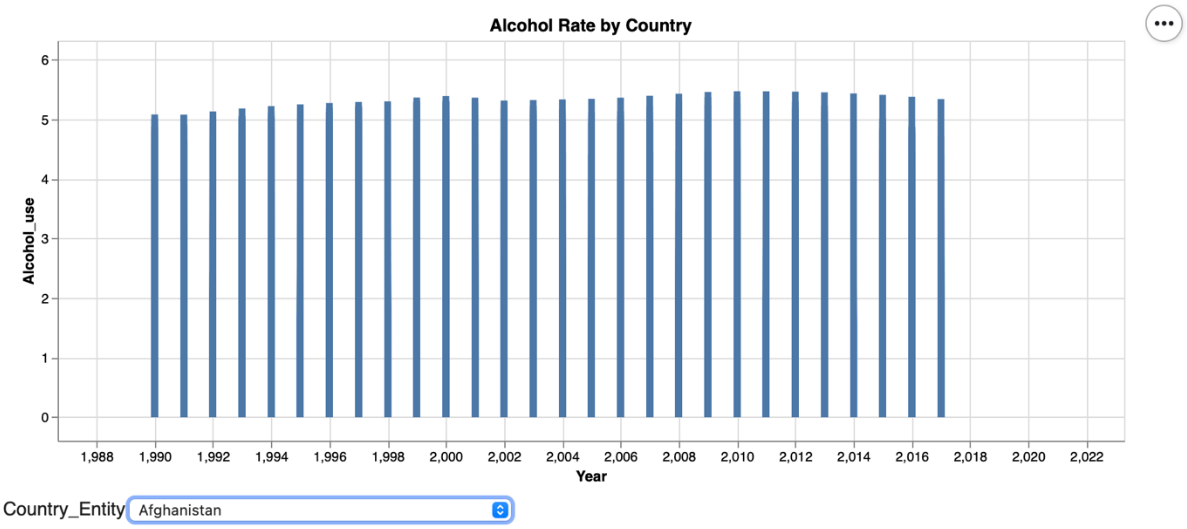


21 to 30 of 6468 entries

Entity	Code	Year	Alcohol_use
Afghanistan	AFG	2010	0.662061905219
Afghanistan	AFG	2011	0.662254250874
Afghanistan	AFG	2012	0.662372139841
Afghanistan	AFG	2013	0.662433436029999
Afghanistan	AFG	2014	0.662446625296
Afghanistan	AFG	2015	0.662276220458
Afghanistan	AFG	2016	0.661850330748
Afghanistan	AFG	2017	0.661217393723
Albania	ALB	1990	1.70946469435999
Albania	ALB	1991	1.70813533017

Show 10 per page

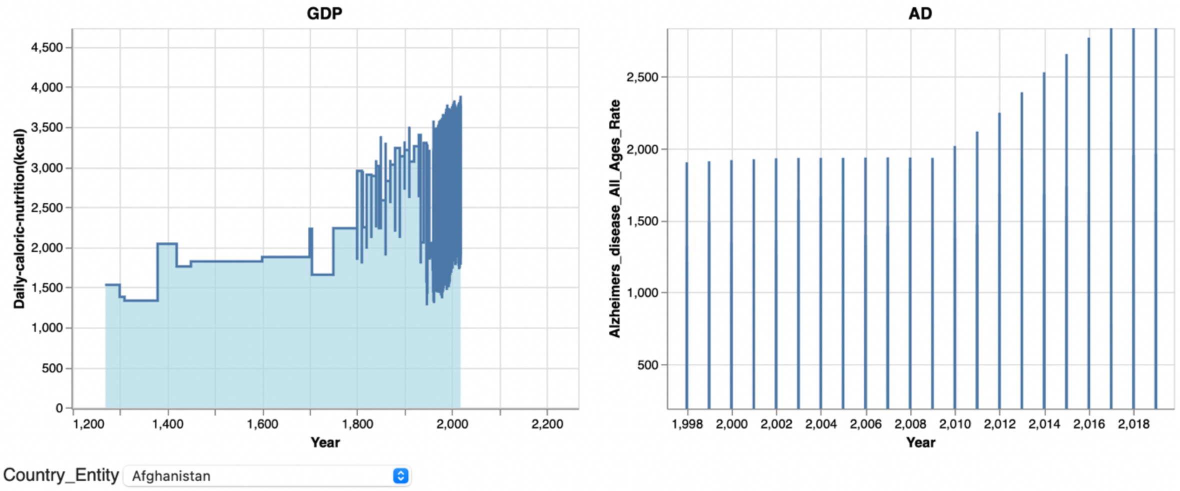
Figure 15 *Interactive Datatable*



Country_Entity Afghanistan

Figure 16 *Interactive Graph*

In the continuation of the analysis, I tried to associate the datasets of these factors with AD and Depression. First of all, I made my analysis of the factors for AD and depression as well. Then I tried to observe the relationship between AD and Depression and factors with graphs. Then I created correlation matrices and heatmaps to observe the consistency of these relationships.



Country_Entity Afghanistan

Figure 17 *Interactive Graph Comparation*

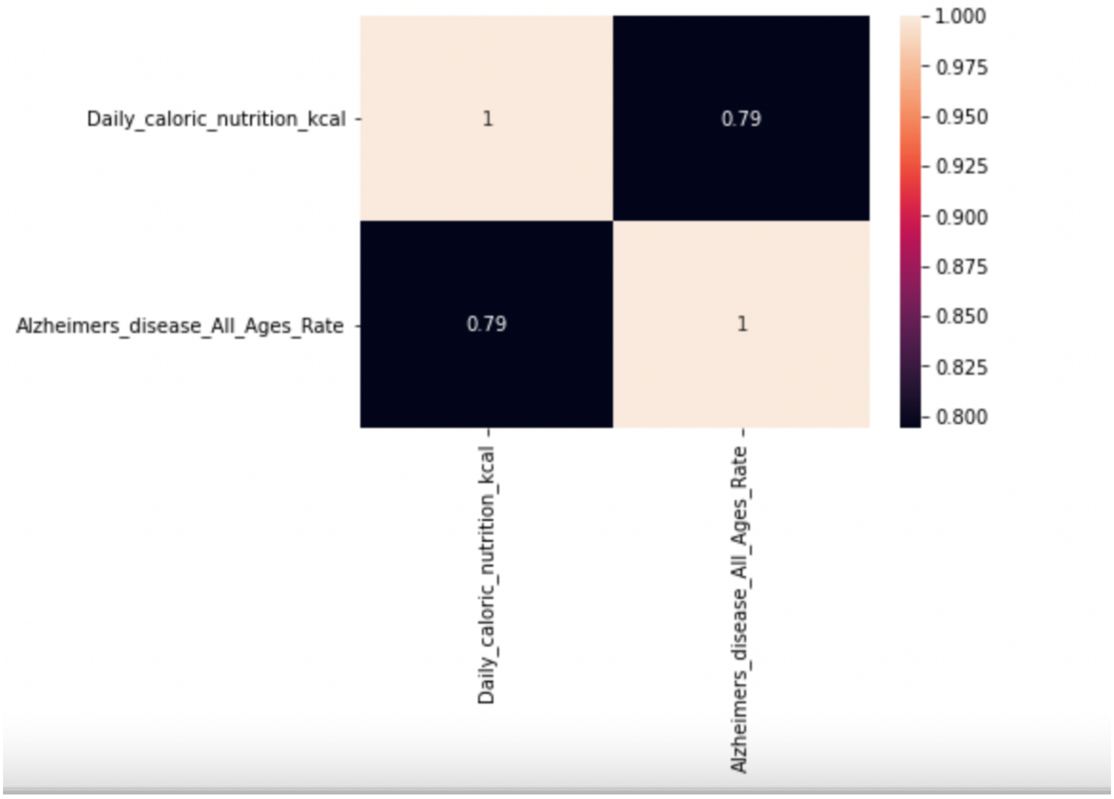


Figure 18 *Heatmap*

Then I made my predictions for AD and Depression and presented them to the user.

```
<matplotlib.legend.Legend at 0x7fe165b917d0>
```

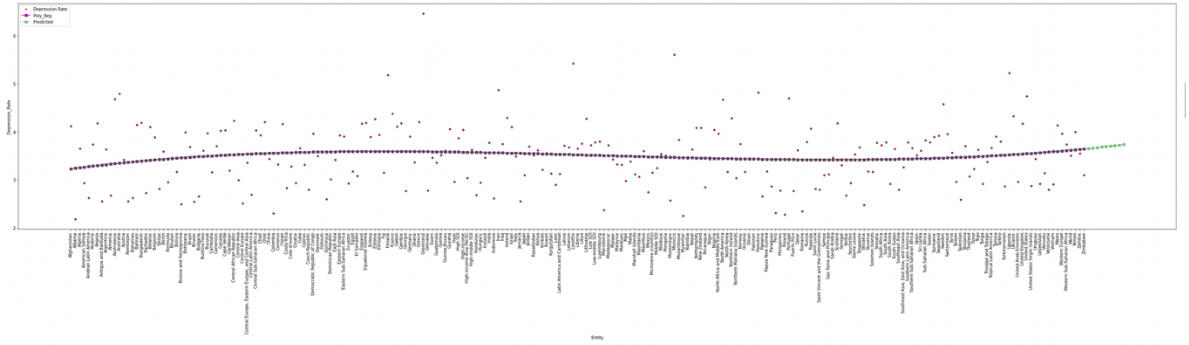


Figure 19 *Regression Table*

In addition, the project offers users the chance to save their work in the user interface.

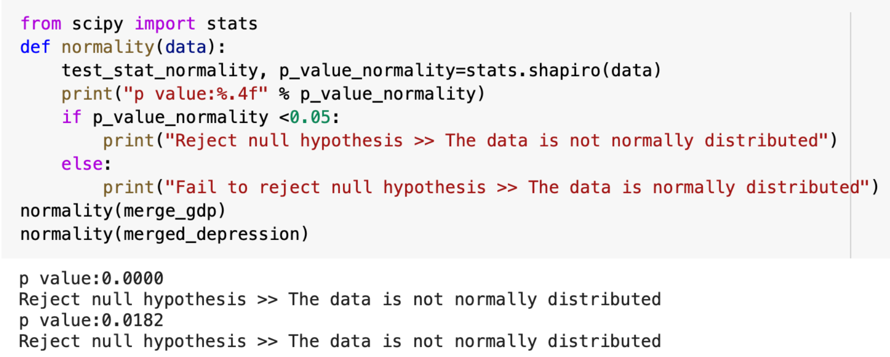
4.3. Hypotheses and Tests

I created my hypotheses based on the datasets I found at the beginning of the project. These:

- GDP and depression have positive relationships.
- There is a positive relationship between alcohol consumption and depression.
- AD and Depression have no relation.
- If a country has a high average education rate, the probability of people having AD decreases
- AD is high in countries with high-calorie consumption.

I came to the conclusions of the hypothesis with the correlation matrices, covariance matrices and heatmaps that I created using the data sets that I generally compared and wanted to find a relationship between. If the correlations between the datasets were high, I concluded my hypotheses according to the result I obtained with the heatmaps. If it is low, I checked its normality with stats-Shapiro tests and then I finalized my hypotheses according to the result I got when I did the t-test. Based on these processes:

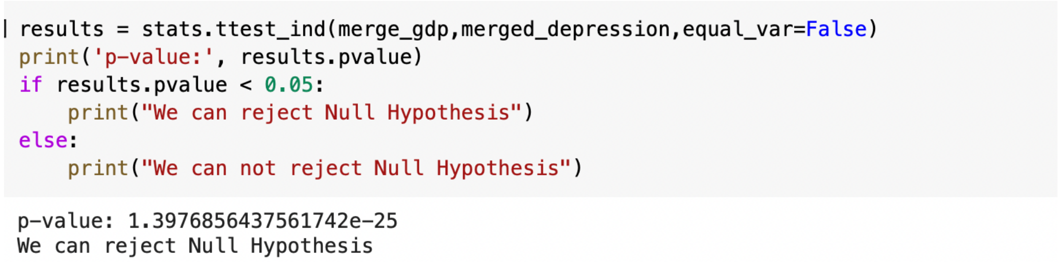
- GDP AND depression have no positive relationships.
- There is no positive relationship between alcohol consumption and depression.
- AD and depression have no identifiable relationship.
- If a country has a high average education rate, then the probability of having AD increases.
- AD is high in countries with high-calorie consumption.
-



```
from scipy import stats
def normality(data):
    test_stat_normality, p_value_normality=stats.shapiro(data)
    print("p value:%.4f" % p_value_normality)
    if p_value_normality <0.05:
        print("Reject null hypothesis >> The data is not normally distributed")
    else:
        print("Fail to reject null hypothesis >> The data is normally distributed")
normality(merge_gdp)
normality(merged_depression)
```

```
p value:0.0000
Reject null hypothesis >> The data is not normally distributed
p value:0.0182
Reject null hypothesis >> The data is not normally distributed
```

Figure 20 *Stats-Shapiro Test*

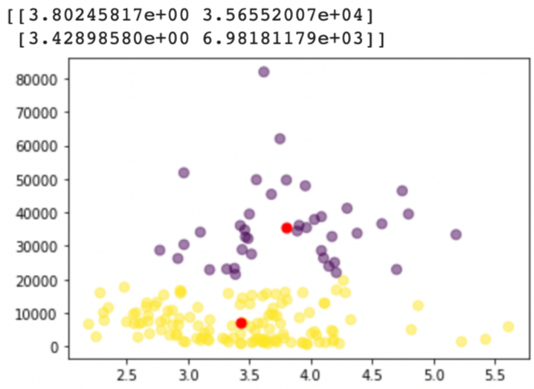


```
results = stats.ttest_ind(merge_gdp,merged_depression,equal_var=False)
print('p-value:', results.pvalue)
if results.pvalue < 0.05:
    print("We can reject Null Hypothesis")
else:
    print("We can not reject Null Hypothesis")
```

```
p-value: 1.3976856437561742e-25
We can reject Null Hypothesis
```

Figure 21 *T-Test*

If I wasn't sure about my hypotheses, I looked at the closeness and distribution of the centers of the datasets by performing kNN classification and tried to execute conclusions.



```
[[3.80245817e+00 3.56552007e+04]
 [3.42898580e+00 6.98181179e+03]]
```

Figure 22 *kNN Classification*

5. Conclusion and Feature Work

The Data Mining Model for Factors of Alzheimer's and Depression is a data mining project created by creating a user interface developed with research-based analysis. In the project, I tried to combine the data analysis studies with the user interface to create a data minig model, and I had a great success. I tested the model with 5 different feedbacks from 5 users. Users were able to analyze and save analysis results by uploading their own datasets to the model. The hypotheses foreseen at the beginning of the project were tested and finalized. It was observed that AS and Depression were correlated with the determined factors. Results from the analyzes:

- There is no strong positive relationship between alcohol consumption and depression.
- There is no strong positive relationship between smoking and depression.
- Females' depression rates are higher than males.
- Depression cannot be defined by age
- There is no strong positive relationship between GDP and depression.
- There is no strong positive relationship between calorie nutrition and depression.
- There is no strong positive relationship between average education and depression.
- There is no strong positive relationship between alcohol consumption and AD.
- There is a nearly strong positive relationship between smoking and AD.
- There is a nearly strong positive relationship between GDP and AD.
- There is a highly positive relationship between calorie nutrition and AD
- There is a nearly positive relationship between average education and AD.

The model is suitable for development. The datasets will be updated on the platform on which the project is presented at regular intervals so that the model remains up to date. In addition, the project can be further developed with more consistent and large data sets. The model can be studied with new data sets. Thus, the Model will also give direction to new analyzes. On the other hand, the user interface can be enhanced and presented in a more observable way that will appeal to the user more.

6. 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. 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
4. 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
5. Gender Differences in Causes of Depression (2001), https://www.tandfonline.com/doi/abs/10.1300/J013v33n03_11
6. 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
7. Food Combination and Alzheimer's Disease Risk: A Protective Diet (2010), https://pubmed.ncbi.nlm.nih.gov/20385883/
8. 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
9. AD Dataset https://ourworldindata.org/mental-health
10. Depression Dataset: https://data.world/vizzup/mental-health-depression-disorder-data
11. Calorie Nutrition Dataset https://ourworldindata.org/calorie-supply-sources
12. Smoking Dataset https://ourworldindata.org/smoking
13. Depression by Age: https://ourworldindata.org/mental-health
14. Depression by Genre: https://ourworldindata.org/mental-health
15. GDP Dataset: https://ourworldindata.org/grapher/gdp-per-capita-worldbank
16. Average Education Dataset: **https://ourworldindata.org/global-education**
17. Alcohol Dataset: https://data.world/vizzup/mental-health-depression-disorder-data