Cyprus University of Technology
Faculty of Communication and Media Studies

Bachelor's Thesis

WEB-BASED DECISION MAKING USING MACHINE LEARNING AND RECOMMENDER SYSTEM: THE CASE OF EXODUS

Adamos Stavrinos

Limassol, May 2018

CYPRUS UNIVERSITY OF TECHNOLOGY

FACULTY OF COMMUNICATION AND MEDIA STUDIES

DEPARTMENT OF COMMUNICATION AND INTERNET STUDIES

Bachelor's Thesis

WEB-BASED DECISION MAKING USING MACHINE LEARNING AND RECOMMENDER SYSTEM: THE CASE OF EXODUS

Adamos Stavrinos

Supervisor

Faculty of Communication and Media Studies, Costas Tziouvas, Special Teaching Staff

Limassol, May 2018

Copyrights

I would like to thank especially my supervisor, Dr. Costas Tziouvas for the support and the knowledge he managed to pass on to me until the end of this research. Furthermore, I would like to thank Dr. Iolie Nicolaidou for helping me choose the correct questions for the questionnaire. Moreover, I would like to give a huge thank you to my family and my girlfriend who showed support and understanding during the difficult situations on this research. Also, I would like to thank my friends as well, who supported me in the whole process.

ABSTRACT

On a daily basis, each one of us is required to decide on a number of issues. Difficulty varies depending by the importance and the options of the decision. As the time passes by, technological advances help millions of users in decision making by creating new methods and tools that users can easily use and utilize. The aim of this study was to create an online platform that assists users in the case of exodus decision. To do so, state of the art technologies will be used including M.E.A.N (MongoDB, ExpressJS, AngularJS, NodeJS) stack and machine learning. Furthermore, the methods used are explained in detail as well as the reason M.E.A.N stack was selected against other available technologies. At the end of this research, the platform implemented was evaluated, providing users with an order list of best matching options for exodus based on the answers a user provided in a preceding questionnaire. In addition, this study evaluates the performance of machine learning in the domain of exodus selection. The research concludes with some results regarding the successful fulfilment of all research question. Also, we will provide some results of how machine learning reacted with the results given.

Keywords: machine learning; recommender systems; web developing; decision making

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

SQL:	Structured Query Language
JSON:	JavaScript Object Notation
HTML:	HyperText Markup Language
CSV:	Comma Separated Values
ARFF:	Attribute-Relation File Format
PHP:	Hypertext Preprocessor
RFID:	Radio Frequency Identification

1 Introduction

World Wide Web is evolving with outstanding pace creating innovative and wonderful tools that are utilized billions of people each day. To be more precise, 3,921,500,000 billion of people as of today are using the World Wide Web (Internet Live Stats, 2018). Each day, those people, are facing with numerus decisions, either with small impact such as "What to eat?", or life changing such as "Should I move to New York or stay at my home country?".

In this study, we are investigating if a web-based platform can help people make easier and more accurate decisions in the case of exodus. To the best of our knowledge, we could not find any similar platform that could help the user make an easy and accurate decision regarding his/her next exodus.

The necessity of this study is of high importance since there are no other examples of such decision-making platforms or similar researches in the past. We will contribute on a more general approach to the literatures about Machine Learning and Recommender Systems. We also want to find out if a web-based decision-making platform with an intuitive interface will have a significant impact on peoples' everyday life. In addition, users will need less time to decide where to go out and there is a great percentage of users that by using the platform will eventually enjoy more their hangout activities. In conclusion, by using the platform users will save money, time, and most importantly have a good time in every exodus.

As we mentioned before, the platform will utilize a questionnaire in which the users will have to answer eight easy questions so that the core system (Machine Learning) will recognize their mood depending on the answers pattern.

By the end of the questionnaire, the platform will suggest (Recommender System) to the users a list of restaurants, bars or clubs so they can enjoy their free time depending on their needs and mood.

The rest of this paper is organized as follows. First, in chapter two, we will state the research questions and some assumptions for this study. In chapter three, we will examine the theoretical background of the tools and methods we are going to use, such as, Machine Learning, Recommender Systems and the collection of M.E.A.N Stack. In chapter four, we will discuss some related work which will help us understand many of

the tools and methods included in this study. Moreover, in chapter five, we analyze the steps of the methodology we are going to follow. Also, in chapter six, we point out some of the ethical issues arising on the platform usability. Additionally, in chapter seven, we explain in great detail every step of the design and implementation of the platform. Finally, in the conclusion chapter, we will demonstrate the results of this study regarding the research question.

2 Research Questions and Assumptions

2.1 Research Questions

The main reason of conducting this research is to create an innovative web-based platform which will support many people in the decision of their exodus. First of all, we need to focus on the questionnaire that will be utilized in order for the system to identify the need of a user. We must make targeted questions which will allow our method, machine learning, to provide accurate results to the users. Thus, the question of this research is “Is it possible to create a platform where machine learning is implemented, in order to give an accurate recommendation to the user?”. In chapter 5 we will discuss in depth how we are going to evaluate our research question.

2.2 Research Assumptions

In the Research Assumptions we will identify two hypothetical scenarios that will be evaluated at the end of the research. In our first hypothetical scenario, we want the platform to help the users on the issue of choosing their next exodus. Also, we want the users to use the platform on a daily basis. In our second hypothetical scenario, we want to find out how accurate was the result suggested by the platform.

3 Theoretical Background

In the current chapter, we are going to discuss the technologies and tools utilized in this research. We chose two tools/technologies for our research which are the Machine Learning and the collection of M.E.A.N. stack. These two tools will provide us with the proper knowledge to accomplish the web-based platform of decision making.

3.1 Machine Learning

Machine learning is a method which machines have the ability to learn mostly by humans. It was firstly introduced in 1959 by Arthur Samuel who was working at that time in IBM Computer Company on a project of Checkers game. He wanted to create an artificial intelligent opponent, which could play Checkers versus a human (Wiederhold & McCarthy, 1992). Machine learning is popular for its artificial intelligence part. This method is used to analyze huge data sets using algorithms and sort them out using labels (Kononenko, 2001). For example, as we can see in the Figure 1, machine learning gets the metrics of a Sepal Length, Sepal Width, Petal Length and Petal Width. As a result, machine learning labels the type of flower with the given class (Analyticskhoj.com, 2015).

	Sepal.Length	Sepal.Width	Petal.Length	Petal.Width	Species
1	5.1	3.5	1.4	0.2	setosa
2	4.9	3.0	1.4	0.2	setosa
3	4.7	3.2	1.3	0.2	setosa
4	4.6	3.1	1.5	0.2	setosa
5	5.0	3.6	1.4	0.2	setosa
6	5.4	3.9	1.7	0.4	setosa
7	4.6	3.4	1.4	0.3	setosa
8	5.0	3.4	1.5	0.2	setosa

Figure 1: Sepal sample labeled by Machine Learning

As mentioned in Domo website (Sparks, 2017), the top 5 companies which use Machine Learning and examine this methodology very well by making a good use of it are: Amazon, Netflix, Google, Salesforce and IBM.

Amazon uses this method for product recommendations, supply chain, forecasting and capacity planning. In addition, another product by Amazon which is famous for its

interaction possibilities is Alexa. Alexa is an AI system understanding human commands such as “What is the forecast outside?” and is capable of answering to these commands. Prime Air is a service also by Amazon which delivers by drone (Unmanned aerial small-scale vehicle) the package ordered through Amazon website to the doorstep of the buyer. Finally, another service by Amazon which depends on Machine Learning is Amazon GO. Amazon GO is a classic store but the plot twist in this is they do not use personnel to run the store. Customers enter the store, make their shopping and leave the store without needing any store personnel. By using the technology of RFID, products get scanned when the customers leave the store and the customer gets charged accordingly on the point.

Another big company which uses Machine Learning is Netflix. Netflix is a digital streaming company which provides entertainment to the user by allowing them to watch Movies, Series, Documentaries and so on. As mentioned in the article of Ms. Sparks, Ioannia Katsavounidis which is a Sr. Research Scientist at Netflix said that, “We can now optimize scene by scene with an almost infinite matrix of possibilities.” which in simple words by the use of Machine Learning it reduces the amount of data that videos are needed to stream and optimizes automatically the video for lower-bandwidth, resulting in an unstoppable view experience for their users.

The next world-renowned company that uses Machine Learning is Google. Google focused on deep learning since they aced their search engine which is beloved by users around the globe. As mentioned in the article of Ms. Sparks, “Google’s Machine Intelligence efforts have focused on deep learning, which involves multiple layers of neural networks—built to simulate human thought processes—that allow Google’s technology to process data more thoroughly.”.

Moreover, Salesforce is a customer relationship management platform which provides the users with applications such as sales, service and marketing in a more simplified way which does not require any special IT knowledge for functioning. Their Machine Learning algorithm provides their users with automatic discovery of relevant insights, predicts customer behavior, recommends actions to follow and automates tasks like logging customer data, capturing sales activity, messaging customers when they are most likely to engage and so on.

Finally, IBM uses Machine Learning called Watson which won Jeopardy in 2011. As the article mentioned, "Watson was designed, however, to consume data at a drastically faster rate than a human can, to learn from it, and augment human capabilities. In the case of cancer treatments, Watson can read half a million medical research papers in 15 seconds and was trained at Memorial Sloan Kettering in New York to be able to suggest diagnoses and treatments to doctors". In general, it can provide a result within milliseconds in contrast to humans, who can take up to years to conclude a result.

As we analyzed above, big companies have adopted this method (Machine Learning) that can be used in an infinite number of scales, providing a result within millions of data with ease and speed.

3.2 Recommender Systems

Recommender Systems is based on the information filtering system which separates a huge data to rating sections. For example, some big companies use recommender systems, such as amazon.com, for recommending items to the users based on their recent activity. Another example of a company that uses this method is youtube.com which suggests videos to its users based on their past views of similar videos or suggests a video based on the content of the users likes, for example gaming, cooking, automobiles channels and so on. It can be used in a lot of areas such as music, books, research articles, online dating and more (Bobadilla, Ortega, Hernando, & Gutiérrez, 2013).

3.3 The Collection of M.E.A.N Stack

M.E.A.N stack is a collection of frameworks creating interactive, fast and reliable websites with the programming language of JavaScript (Kadam, Chaudhari, Patil, & Chavhan, 2017).

3.3.1 MongoDB

Mongo Database was released in 2009. By the end of 2015, Mongo Database was the most popular document-oriented NoSQL database. Mongo Database is still having the upper hand in the databases field because it can arrange unstructured data like documents, e-mails, videos and music. With the NoSQL methodology we mean, non-

relational databases and the data is not stored in tables, schema is not static, and it has a more simplistic data model rather than an old structured SQL database, such as MySQL. Data can be stored in a binary JSON format which makes it easier to transfer data between client and server.

In a study made by Gyorodi, Gyorodi, Pecherle and Olah, (2015) they comparing MySQL and MongoDB by going through several tests, in order to answer the question of which one is better to use and why. For the testing process they used 5000 rows for each table. The tables where forums, sub forums, discussions and comments. Every table included 5000 entries.

First test was "insert" the data as mentioned above to the databases. As we can see in Figure 2, MySQL needed 1010 seconds rather than MongoDB where only 3.3331 seconds were needed. MongoDB reacted faster in inserting data and performed much better than MySQL.

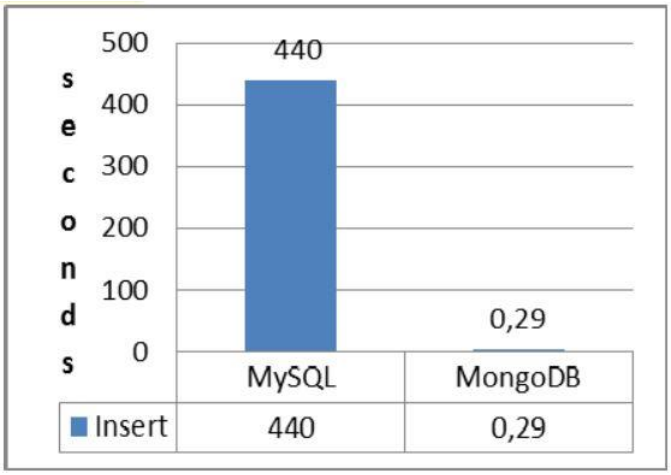

Figure 2: MySQL and MongoDB time comparison for Insert Command

In the second test, they wanted to test how well those two different databases structures will handle the "select" command of a whole table/collection. MySQL completed the action within 0.6478 seconds. On the other hand, as we can see in the Figure 3, MongoDB has completed the task in 0.0052 seconds which is significantly better.

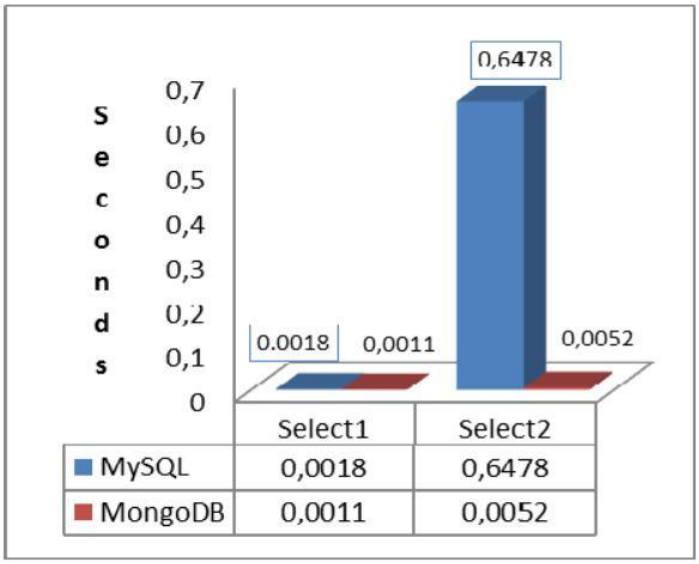

Figure 3: MySQL and MongoDB time comparison for Insert Command

In the third test they wanted to observe how well the databases will handle "update" command for a table/collection. In Figure 4 we can acknowledge the significant difference between MongoDB and MySQL. MongoDB completed the "update" command within 0.0021 seconds rather than MySQL which completed the task within 0.0987 seconds.

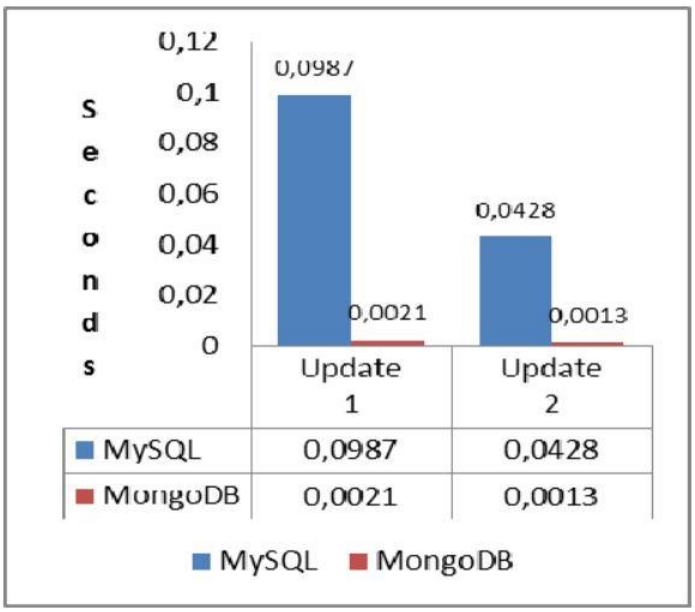

Figure 4: MySQL and MongoDB time comparison for Update Command

For the final test, they wanted to perceive the difference for the “delete” command. In Figure 5, we can understand the difference of the results which MySQL finished the task with 0.8231 seconds and MongoDB completed the task within 0.0064 seconds.

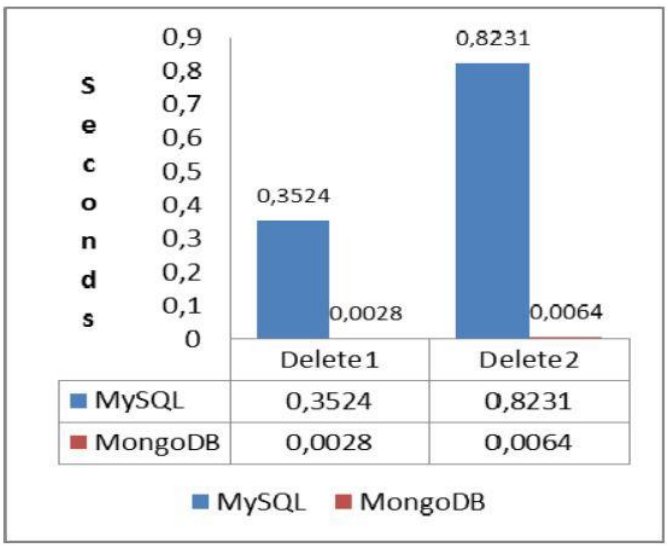

Figure 5: MySQL and MongoDB time comparison for Delete Command

In conclusion, MongoDB performed much better with lower execution times than MySQL in all the above tests. We can say with confidence that MongoDB provides flexibility, power, speed within the package. Moreover, in a larger scale company for example eBay, Foursquare, Google and so on, where the users will simultaneously send and receive similar commands to the database, this execution times will perform a vital importance to the user experience and the flexibility MongoDB offers to the developer.

3.3.2 ExpressJS

ExpressJS is a framework created on 2009 by Node Foundation and it is used to create web applications and API’s under the rooftop of Node, which helps to deliver the web application to a computer screen and a smartphone screen with ease (Express.js, 2018). With Node by its side, it can create an outstanding website using only two programming languages, JavaScript and HTML. ExpressJS is adding the extra layer above Node which makes it an amazing interactive website for web applications.

As mentioned in their legitimate ExpressJS website (Express.js, 2018), some big companies are using ExpressJS. Some of those companies who are using ExpressJS is IBM, QuizUp, Uber and FOX Sports.

3.3.3 AngularJS

AngularJS is a framework developed by Google in 2009. As mentioned in the research of Jain, Mangal, and Mehta (2015), the benefits of using AngularJS are many, starting with Integration with existing applications. This means you can add more in your basic index.html without any obstacles of incompatibility to be faced.

Simplicity is one of the main factors for using AngularJS. Scripts can be run locally without using any web server to compile. This can be found very useful for people who create quick mockups “on the go” on their machine.

Another basic factor is Extensibility of this framework. A web developer can use this benefit to use directives as AngularJS, allowing the creation of custom elements and attributes which can extend an HTML file.

Other worth mentioning features of AngularJS are the two-way data-binding, templates and MVC (Model View Controller).

Two-way data-binding is the ability to “call” a function within the HTML tags by a script implemented inside the HTML. An example is given in Figure 6 where we can understand what exactly two-way data-binding is.

```
01<!doctype html>
02<html ng-app>
03  <head>
04    <script src="http://code.angularjs.org/angular-
05.1.0.0rc10.min.js"></script>
06  </head>
07  <body>
08    <div>
09      <label>Name:</label>
10      <input type="text" ng-model="yourName"
placeholder="Enter a name here">
11      <hr>
12      <h1>Hello, {{yourName}}!</h1>
13    </div>
14  </body>
</html>
```

Figure 6: Data-Binding sample of AngularJS

The highlighted text is the way data-binding works. It receives the input from ng-model and transfers it to the brackets yourName. The easiness of this, can save us enough time and when the project is getting larger code wise, this is very helpful.

When it comes to Templates, AngularJS provides the developer with the easy use of a function within HTML tags for multiple times. For example, in Figure 7, we can see an algorithm where a collection of images is shown in the browser. The author of the paper, used the function ng-controller=”AlbumCtrl”, ng-repeat and ng-src. The author uses those functions in order to show how the pictures are presented.

```
<script>
  function AlbumCtrl($scope) {
    scope.images = [{
      "thumbnail": "img/image_01.png",
      "description": "Image 01 description"
    }, {
      "thumbnail": "img/image_02.png",
      "description": "Image 02 description"
    }, {
      "thumbnail": "img/image_03.png",
      "description": "Image 03 description"
    }, {
      "thumbnail": "img/image_04.png",
      "description": "Image 04 description"
    }, {
      "thumbnail": "img/image_05.png",
      "description": "Image 05 description"
    }];
  }
</script>
<div ng-controller="AlbumCtrl">
  <ul>
    <li ng-repeat="image in images">
      <img ng-src="{{image.thumbnail}}" alt="{{image.description}}">
    </li>
  </ul>
</div>
```

Figure 7: AngularJS implementation of Templates

MVC also known as Model View Controller, is how AngularJS builds client-side web applications. The data inside an application can also be called models. In general, models are old JavaScript objects where there is no need to inherit from the framework classes or use special methods or sets to access it. When we are working with old javascript called “vanilla” javascript, meaning that there are no libraries implemented in

the code, and when using the model there is no need to remove application boilerplate code.

View is when the JavaScript code is compiled and parsed to the HTML to include rendered markup and bindings. In a more general scale the view is handling the layout of how your webpage will show up to the device's screens.

The main purpose of the controller is the initial state and augmenting with the command $scope, therefore, it controls the behavior.

Additionally, AngularJS have made an outstanding entrance to the web development and web design, with reaching hundreds of thousands of views worldwide in programming genre websites. As we can see from Figure 8, AngularJS reached 27.2 thousand stars on Github which is among the best websites for programming-based repositories. Backbone.js on the other hand, has 18.8 thousand stars and Ember.js has 11 thousand stars.

Metric	AngularJS	Backbone.js	Ember.js
Stars on Github	27.2k	18.8k	11k
Third-Party Modules	800 ngmodules	236 backplugs	21 emberaddons
StackOverflow Questions	49.5k	15.9k	11.2k
YouTube Results	~75k	~16k	~6k
GitHub Contributors	928	230	393
Chrome Extension Users	150k	7k	38.3k

Figure 8: Popularity comparison across different platforms of AngularJS, Backbone.js and Ember.js

Furthermore, Third-Party Modules means, a code created by a third party and neither you nor the company/organization created the programming language created the

module. At the stage of AngularJS, 800 was recorded of third-party modules created. The other frameworks had less significant modules rather than AngularJS.

They also recorded the Stack Overflow Questions. Stack Overflow is again one of the biggest Q&A programming websites around the globe. People asked a total of 49.5 thousand questions about AngularJS rather than Backbone.js and Ember.js. Backbone.js had reached a total of 15.9 thousand questions and Ember.js 11.2 thousand.

Moreover, another metric they recorded was YouTube results. Approximately 75 thousand results came up for AngularJS. Backbone.js and Ember.js combined had reached the approximate number of 24 thousand searches.

Also, AngularJS had 928 contributors on Github, showing that users were really keen in contributing to this framework probably because they appreciated its functions. On the other hand, Backbone.js and Ember.js did not have as many contributors as AngularJS. Backbone.js had 230 Contributors and Ember.js had 393 Contributors.

Together with the other frameworks they combine an outstanding fast front end development website. It extends an HTML file with attributes <script>. In other words, Angular adds an extra layer to make an even more beautiful and interactive website in the basic HTML code.

3.3.4 NodeJS

NodeJS is written in JavaScript language and is an open-source, cross-platform environment used mostly for server platforms. It has significant differences between PHP/Apache stack which is written in PHP language and it is also a server sided. In the conclusion of a research made by Chaniotis, Kyriakou and Tselikas (2015) where they compared the performance between Apache, nginx and NodeJS found that, Apache have been by far the poorest in performance tests regarding NodeJS and nginx.

3.4 Waikato Environment of Knowledge Analysis

W.E.K.A is a state-of-the-art data mining program, coded with the programming language Java and has been developed in 1992 by the University of Waikato. On Source Forge W.E.K.A has been downloaded from 2000 till today 9.6 million times as we can see in Figure 9 (SourceForge, 2018). W.E.K.A is moderately popular program in the

past 5 years since Machine Learning and data mining has become one of the most essentials of knowledge. It is an open source program completely free of charge with main purpose of analyzing big data and the use of the machine learning.

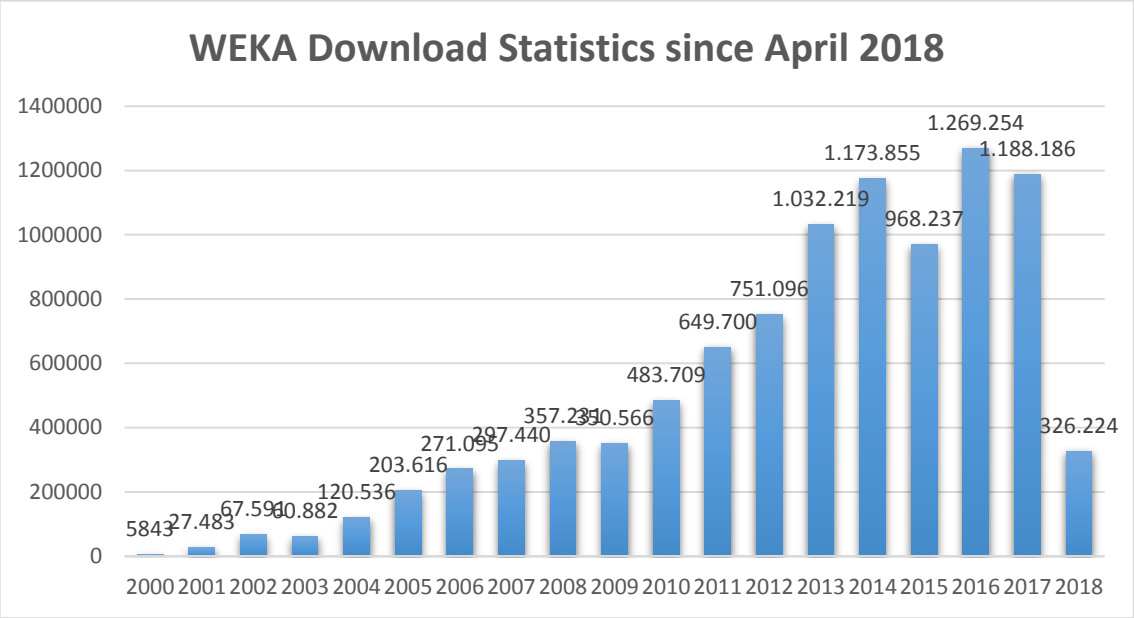

Figure 9: W.E.K.A download statistics since April 2018

3.4.1 Graphical User Interfaces

W.E.K.A has several graphical user interfaces which enable an easy access the features of the program. The main window that comes up when we run W.E.K.A, is a simple layout with 5 options and those are “Explorer”, “Experimenter”, “KnowledgeFlow”, “Workbench” and “Simple CLI”. For this purpose of the study we are going to use the “Explorer” option since we need to analyze the data. As we can see in Figure 10, “Explorer” has several tabs of different functionalities of the program. The first tab is called “Preprocess” and the main purpose of this feature is to load datasets and transform them using the supportive tools which are called “Filters”. The supported formats that W.E.K.A can handle is ARFF format (W.E.K.A Format), LibSVM’s format, CSV format and C4.5’s format.

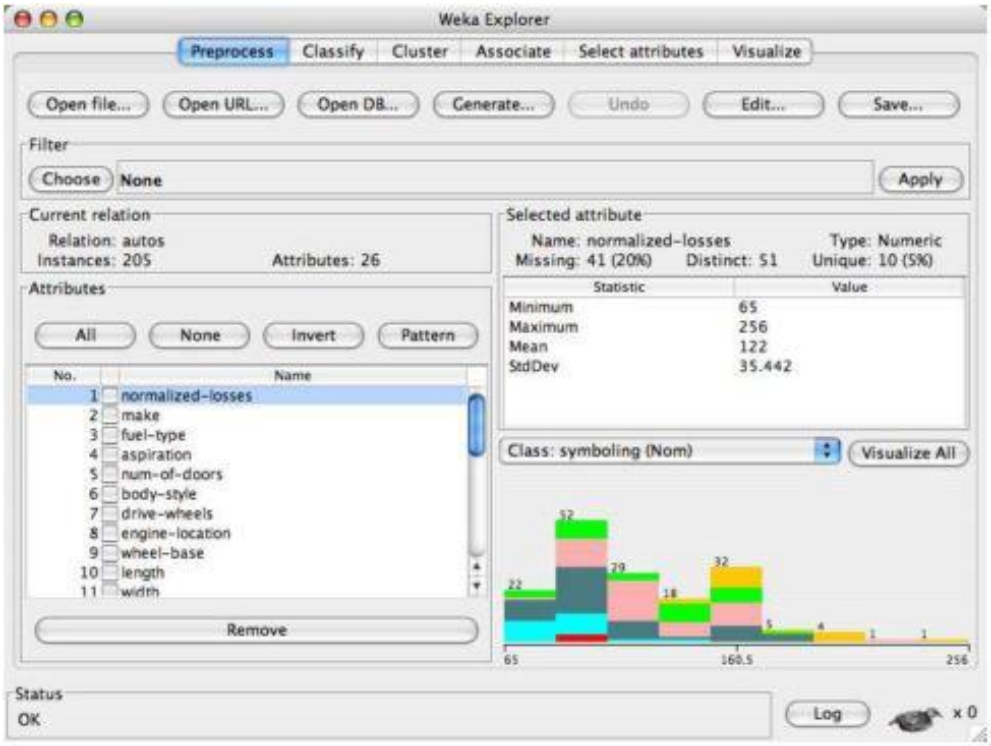

Figure 10: Graphical User Interface of W.E.K.A program

The second tab is called “Classify”. Classifications are used for supervised learning. That means, is most frequently used for a dataset of prior knowledge of classes and it classifies new sample into the known classes. Furthermore, at this tab we can visualize the prediction of errors in scatter plots and other types of visualizing methods.

The third tab is named “Cluster”. Clustering on the other hand is an unsupervised method used for group up objects. It discovers if there is a relationship between objects and by the patters of data it learned, it suggests a group to be added on. We have also the visualization function if possible by the dataset of the clustering structure.

Finally, the last tab in the Explorer is named “Visualize” which provides numerous plots to visualize the data loaded from the preprocessing tab.

3.4.2 Classification Algorithms

W.E.K.A contains pre-installed learning algorithms for easy analyzing big data both supervised and un-supervised such as Bayesian Logistic Regression, Best-First Decision Tree, Decision Table Naïve Bayes Hybrid, Discriminative Multinomial Naive Bayes, Functional Trees, Gaussian Processes, Simple CART, Variants of AODE, Wrapper

Classifier, Nested Dichotomies, Dagging, Rotation Forest, CLOPE Clusterer and Sequential Information Bottleneck Clusterer.

3.4.3 Preprocessing Filters

Preprocessing filters are the filters added to the dataset loaded before continuing to any other functionality of the program. Some examples of those filters are Add Classification, Add ID, Add Values, Attribute reorder, Interquartile Range, Kernel Filter, Numeric Cleaner, Numeric to nominal, Partitioned Multi-Filter, Propositional to Multi-Instance and Vice Versa, Random Subset, RELAGGS, Reservoir Sample, Subset by Expression and Wavelet.

4 Related Work

In this chapter, we are going to discuss about related work that have been made in the past regarding machine learning and M.E.A.N. stack technologies. Here, we present, some other studies, which are going to reinforce our research.

One important study is a research paper by Katakis I, Tsapatsoulis N, Triga V, Tziouvas C and Mendez F. (2012). This research captured our attention due to it similarity to the proposed research. They made possible to create a VAA system which stands for Voting Advice Applications. VAAs are survey platforms consisting of several questions and at the end, automatically by using the questions answered, suggests to the users the political party to which they should give their vote. The study is using the machine learning technology for clustering data to the system and recommend to the user a suggestion regarding their next vote.

The second research to catch our eye was by Wang and Summers (2012). In this study we can get familiar with what is machine learning and how it can help us. Authors have studied how machine learning can diagnose a patient with cancer or any other illness by the metrics gained. As we can understand, machine learning is everywhere nowadays, providing to the users an accurate result. Shijun and Ronald in their study, concluded that "Machine learning will be a critical component of advanced software systems for radiology and is likely to have wider and wider application in the near future."

Another helpful research was by Bobadilla et al (2013). This study will introduce us to the knowledge of what a recommender system is, and the methods that it is capable of carrying out. The conclusion from their study showed us that, recommendation tools are a very useful method since the beginning of the web. Furthermore, in recent days, Web 3.0 has become more intelligent and more accurate to the matter of user suggestion. As a matter of fact, recommender systems can analyze any behavior from a user and suggests to them a better idea or a result.

Research number four was by authors Brozovsky and Petricek (2007). In their study, they examined how recommender systems work inside an online dating service. One important aspect of their study is that they reached deep into the field of the algorithms of recommender system which will help us comprehend even more about this method. As they mention, "User-User Algorithm is one of the best-known collaborative filtering

algorithms. When predicting ratings for active user “a”, the user database initially searches for users with similar ratings vectors to the user “a” – “neighbors””. By this saying, we understand that when a user number two has 5 out of 6 answers identical with user number one, both will be added as neighbors, because the recommender system will rank user number two as similar to user number one based on their answers. Similar task will be added to our research platform, so that when the machine learning will identify the answer flow, it will eventually arrange them to a “ranking” system.

Research number five concerns about the programming aspect of this study. The research was made by Kadam et al. (2017). The authors of this article have been studying the pros and cons of building a website based in the M.E.A.N. stack framework. They analyzed very well each aspect of the frameworks (MongoDB, ExpressJS, AngularJS, NodeJS) and how they can be integrated. Furthermore, they analyzed the system architecture and what is the minimum specifications, so a client can “run” a fully created website by M.E.A.N. stack. As we can see in Figure 11, the client makes a request to the website and the website is connected in the NodeJS Server. ExpressJS will process the requests and come with a result. The result will then go to the MongoDB to get the actual result and then return from the ExpressJS section to the displaying response and finally to the client.

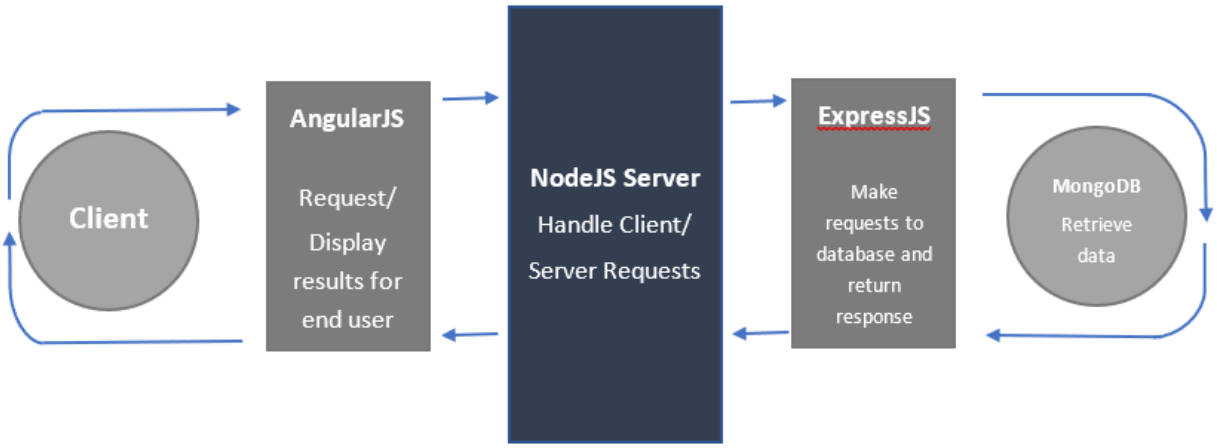

Figure 11: Process of M.E.A.N stack

5 Methodology

For the research part of this study we are going to use a quantitative research method because we are going to create a web-based application. Moreover, our study can be considered as an experimental research due to its high research and development aspect. We will create a questionnaire which will be imbedded in the website so that we can collect users' answers, which will then be used by the machine learning approach for training. The reason we are doing this is, as we already mentioned, "to teach the machine learning how to learn and understand their mood", because each user who is completing the questionnaire, the machine learning will collect their answers, and appropriately will label their sequence of answers to a category of bars or restaurants.

The questionnaire consists of nine simple questions. A more analytical explanation of why we attached the current questions is described in the following table:

Table 1: Questionnaire

Question ID	Question Body	Options	Purpose of Use
Q1	Age	Open Field Answer	To know the exact age of the users
Q2	In which town do you live right now?	1. Limassol 2. Nicosia 3. Paphos 4. Larnaca 5. Famagusta	The reason for this question is to know in which town machine learning will suggest a place.
Q3	How do you feel today in general?	1. Sad 2. Bored 3. Happy 4. Angry 5. Frustrated 6. Calm 7. Flirty	In this way we will understand the mood of the users very easily.
Q4	What would you like the most?	1. Coffee 2. Food 3. Drink	In this question we direct the machine learning towards the users' needs in order to adapt the suggestions accordingly.
Q5	How much money do you want to spend the most?	1. Less than 5 euros 2. Between 5 and 10 euros 3. Between 10 to 30 euros 4. Between 30 to 50 euros 5. More than 50 euros	This question is about the budget that the users want to spend during their exodus. The platform can again adapt the suggestions accordingly.

Q6	What music do you like the most?	1. Rock 2. Acoustic 3. R&B 4. Pop 5. Popular 6. House 7. Blues 8. Classic 9. Greek 10. Anything	By this question we will get familiar with the genre of music the users like, so the machine learning can suggest a place where the music fits with their liking
Q7	How far do you want to go?	1. Walking Distance 2. Drive less than 10 Km 3. Drive somewhere between 10 – 40 Km 4. Drive somewhere between 40 – 60 Km 5. Drive more than 60 Km	The distance users would like to travel to reach their destination.
Q8	What style do you mostly prefer?	1. Classy 2. Cozy 3. Chill 4. Loud 5. Casual 6. Family 7. Lounge	By this question we will understand what is the style that the users like the most in a restaurant or bar, so we can make suggestions accordingly.
Q9	Where do you believe this sequence of questions you just answered should be categorized?	1. Cheapest Priced for Coffee 2. Reasonably Priced for Coffee 3. Expensive Priced for Coffee 4. Cheapest Priced for Food 5. Reasonably Priced for Food 6. Expensive Priced for Food 7. Cheapest Priced for Drink 8. Reasonably Priced for Drink 9. Expensive Priced for Drink	This final question will determine the category for the sequence of answers selected by the users. Users will have to select one of the 9 options in order to label the sequence of answers they just submitted.

Our target is to have 150 questionnaires completed, regardless of age and gender, so we can acquire representative feedback about the platform and also to provide enough data to the machine learning to have a better performance and of accuracy for each

suggestion. In addition, the time frame to complete the questionnaire will be from February 2017 until the end of April 2017 on our online platform in Cyprus. The users can answer the questionnaire in order to get a suggested result for their next exodus.

Moreover, we are going to analyze our data using frequency tables which will give us a better insight of the extend of correct functioning of the platform. The program we are going to use for creating models and analyzing the classification process is W.E.K.A since it is one of the most comprehensive and easy to use Machine Learning programs, that can analyze a large sum of data and label them with the sequence of the answers to a category. At a later stage, this will help us to create a model and implement the model inside the platform. This will be used to suggest a recommendation to the user at the end of the questionnaire.

Finally, the evaluation of our research will be made by the users a few days after the platform has suggested a place for exodus. Users will be prompted to leave a feedback in order to express their satisfaction about the suggestion of exodus received depending on their judgment. Using this method, we will understand if the platform has met the ultimate purpose of this study which was to cater to user's needs and facilitate the important question of 'Where should we go tonight?'.

6 Ethical Issues

Our ethics for this study is to not break any boundaries of identity between the participants and the questionnaire. Every detail of each participant will be fully anonymous and the platform by any means will not expose any personal data of the participants. Also, we do not set boundaries between the age. We believe that this platform should be open to any user without consideration. But of course, there will be a checking system for users under 18 years old. Thus, when a user is under 18, some available options will be blocked in order to not suggest an adult place such as bars and clubs to the user, since in Cyprus it is illegal to consume alcohol under 18 years of age.

7 Design and Implementation of Platform

The design of the platform has been made with the mindset of easy usability for the user and minimum time consumption on the platform, in order to be able to receive results as fast as possible.

7.1 System Requirements

Firstly, and most importantly to this research was the understanding of the platform by the user since it should be easy to use and error-free providing a pleasurable use of the platform.

As we mentioned before, usability is one of the key requirements of the platform. We want users to become familiar with the platform as fast as possible and by achieving that we manage to create a user-friendly platform, easy to use and not time consuming in order to keep the attention and the excitement.

Furthermore, using the M.E.A.N. stack, along with a responsive design, we are provided compatibility to both computer and smartphone, making the platform accessible on every device using internet no matter what their screen size is.

In addition, we had used a M.E.A.N stack boilerplate. M.E.A.N stack boilerplate is a package that have the most fundamental functionality in order to run a basic website. The template although was very simplistic so we needed to make some twitching to proceed to the final step of presenting the website to our users. Website template had only very limited content such as text and a button in the middle. Menu was accessible also with the options of Sign-up and Sign-in.

Moreover, Cyprus University of Technology have provided us with an infrastructure, so we can run and test the platform while we were developing it. They had also provided us with the port “8081”. The domain name they hand over to us was vaa.cut.ac.cy. In conclusion, a user to access the website, had to type the Uniform Resource Locator (URL) vaa.cut.ac.cy:8081. This was not very helpful since the users found the fact that, they needed to type numbers in the link which was moderately complicated and confusing.

Also, we access to the platform is granted using a username and a password on the server of Communication and Internet Studies, so we could install all the proper tools we were going to need by remote connection to the server.

After that, we had to understand how the file structure of M.E.A.N. stack worked, because M.E.A.N. stack provides several services as we mentioned before. Each service had their own small files where all combined could expand the development of the website. We studied the documentation of the file structure within the official website of M.E.A.N. stack (MEANJS, 2017).

By the time we learned how the file structure works, we started designing and developing the website following the needs and expectations of the platform. We had removed any text, buttons and menus that were irrelevant to this project and formulated only basic needs a user might have in order to go through the website in an intuitive manner.

7.2 User Accounts Permissions

There are two user levels assigned to the platform for its use with each user level has different privileges to the platform.

First level is the “Guest” level where users can access the home page of the platform and they can also access the “About us” section, log-in window and sign-up window. “Guest” users do not have access to the questionnaire as it is a must to sign in with an account.

The secondary user level is “User” and with the access of that level, users can enjoy the full experience of the platform. It is required to create an account with several required fields that have to be added. For example, to create an account in the platform you must fill your name, last name, an existing email and a password containing at least 8 characters or numbers.

A user has to create an account in the website in order to proceed to the next step which is answering the questionnaire and get a recommendation depending on the provided answers. Furthermore, we needed this kind of restriction because we wanted to get a legitimate result, and by achieving that, we have to make the user sign up and provide the fields of answers we mentioned above. This has a negative side since the users had

to spend extra time to create an account in order to use this feature of the platform. Also, users might be feeling scared to provide us with such information regarding their personality.

We used the OWASP Password Strength Test module which has been created by NowSecure to provide the security needed by the user who wants to create a password on a website. Moreover, according to this module website (OWASP, 2017), some securities approaches have been implemented to provide to the user an outstanding combination of password strength. The default settings for this module to accept the password was to have an Uppercase letter, a Lowercase letter, at least one number, at least one symbol such as a dot or a backslash, password must not contain the same word twice and lastly, password must not be a phrase for example, "iamflying".

As we can understand, there are a lot of requirements to create a password which can result in the user losing their interest. Consequently, in order to decrease the hassle and the time of creating such a strong password, we set the requirements to create a password to only eight characters containing either numbers, letters or both.

7.3 Questionnaire

The structure and the design of the questionnaire had to be simple and easy to understand the questions. The 9 simple questions in Pre-Alpha version made it helpful for the users and not time consuming. We had used a JavaScript library from SurveyJS.io. SurveyJS is a library where they create different variations of interactive surveys coded with programming language JavaScript. They provide this service with a free Creative Commons License for non-commercial usage and they also offer a paid license for commercial applications.

Furthermore, we used the "Radio Group" survey because at this stage we want the user to be able to select only one answer, so it is easier to the machine learning to deliver a more accurate result at the end of the process.

For the first question, as we mentioned above in Table 1, it asks the user to import their age. We believed that is better to make this question as a "String" type as shown in Figure 12, so the users can implement their exact age with numbers rather than choosing a radial decision such as "18-25" or "26-35" to be more accurate as a result, for the

reason that, there is a small possibility that the users who are 18 years old, might not be willing to go to a place where 25 years old's are hanging out. After the user typed their age, we added an algorithm which transforms the "String" to an "Integer" so it can be added to the MongoDB as an Integer.

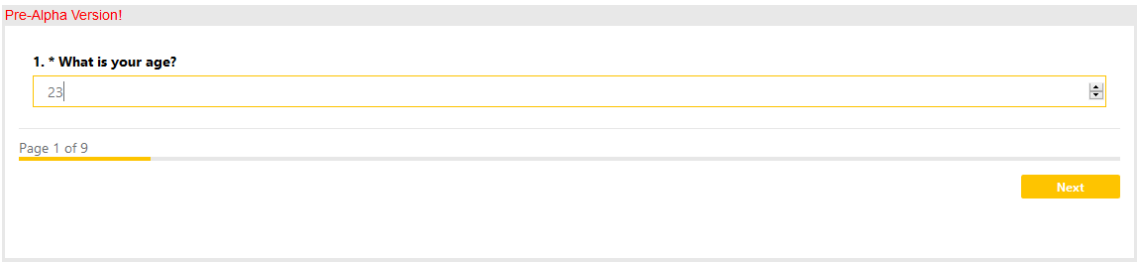

Figure 12: Question number 1 as seen on ExoDecision platform

Moreover, the other 8 questions are radial buttons types. As we can clearly see in Figure 13, the question "In which town do you live" are followed by radial buttons which the options are the same as we mentioned before in Table 1. Each radial button equals to an integer. The identical integer for the option 1 (Limassol) is the number 1. That means, in the MongoDB, will be implemented as number "1" and not as "Limassol". As we mentioned before, numbers are much easier for the machine learning to comprehend by adding integers rather than strings. Moreover, we can manipulate the data to our needs with ease. However, the default option as we can clearly see in Figure 13, is Limassol. For that matter, we made Limassol as the default and only option since it was impossible with the strict time limits we had to collect the data. This is why all other options are grayed out and users cannot select any other option.

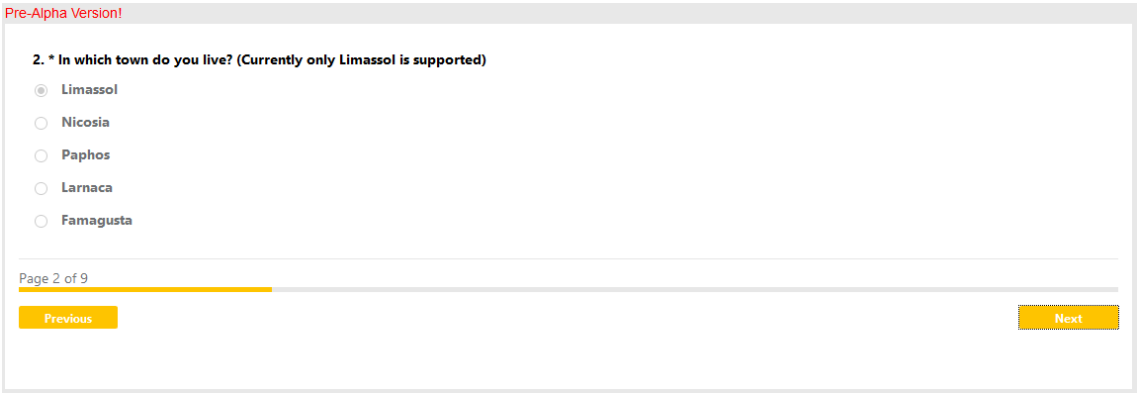

Figure 13: Question number 2 as seen on ExoDecision platform

Additionally, in the question number three, is keeping the same scheme as question number two. Buttons with options arranged with integers. Evidently, in Figure 14 we can acknowledge the design of the question and the purpose of it. The question number three, helps the machine learning to understand the exact mood of the user in order to suggest at the final result a place depending on their mood.

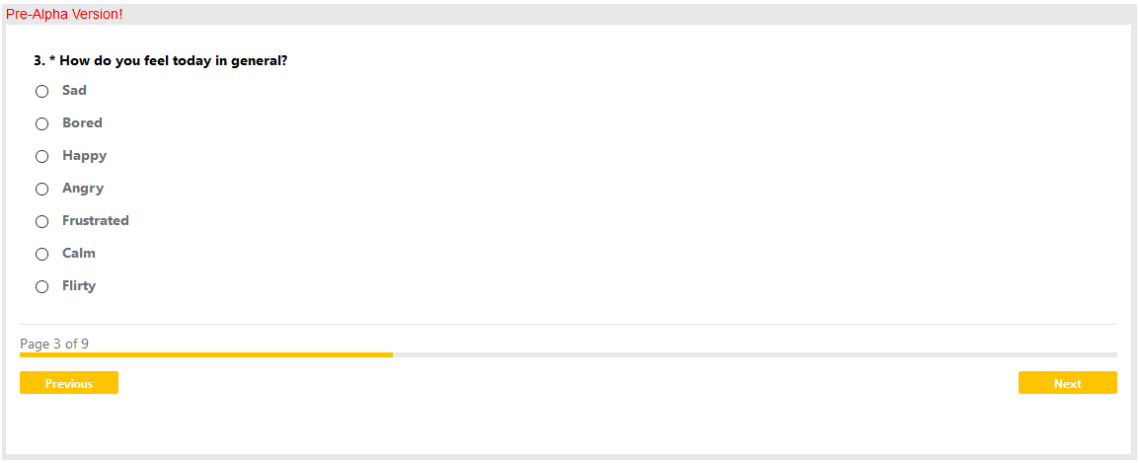

Figure 14: Question number 3 as seen on ExoDecision platform

Question number four, identifies the type of exodus the user would like to be placed on. At that point, machine learning, can understand the type of places the user wants to go and so on, the algorithm, is placing the cluster nearer to the type the user pick. In Figure 15, we can experience the design of the question as is implemented in the platform.

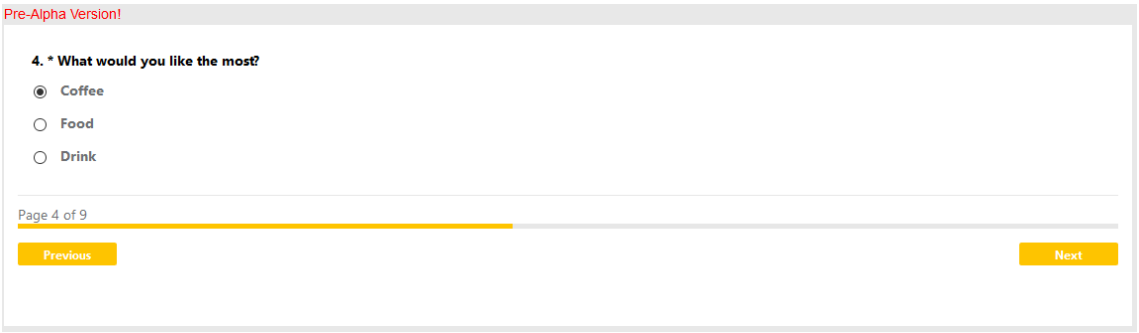

Figure 15: Question number 4 as seen on ExoDecision platform

Likewise, question number five which is in Figure 16, indicates an average on how much money the user wants to spend at their next exodus. By this question, machine learning expands or limits the horizons of the options available to the platform. For example, if a user wants to spend “Less than 5 euros”, the range of options are affected

due to money are quite less. Moreover, if a user chooses the option “More than 50 euros”, the range of results will be much greater, and the platform automatically will choose the right place where the prices are quite high.

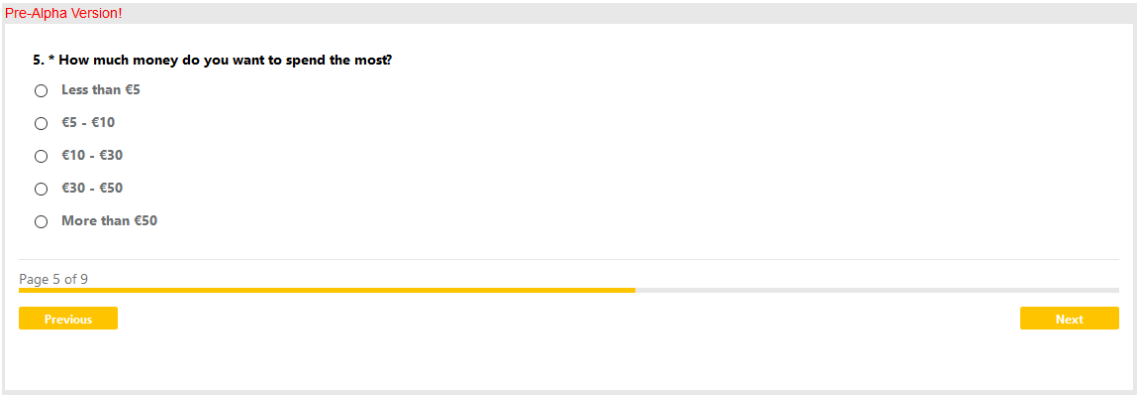

Figure 16: Question number 5 as seen on ExoDecision platform

Similarly, question number six as can be seen in Figure 17, will identify the music preference of the user depending on their likings so in order to aid the machine learning to place the cluster near the preferred music type. We tried to provide to the user a lot of options in this matter and that’s why we include 9 types of music and a 10th option which is “Anything” so we can cover the range of music genres.

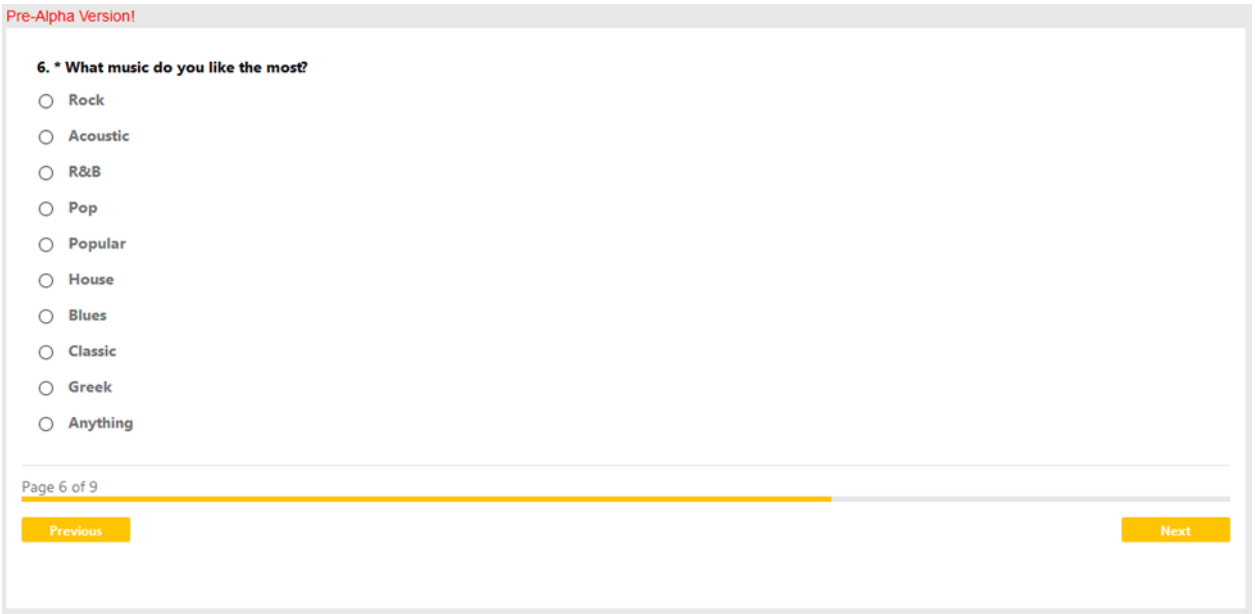

Figure 17: Question number 6 as seen on ExoDecision platform

Also, in the upcoming question number seven, users must select an option depending on the distance they want to travel (How far away the place they want it to be). As we can see in the Figure 18, we cover a decent amount of distance in relation to the size of Cyprus.

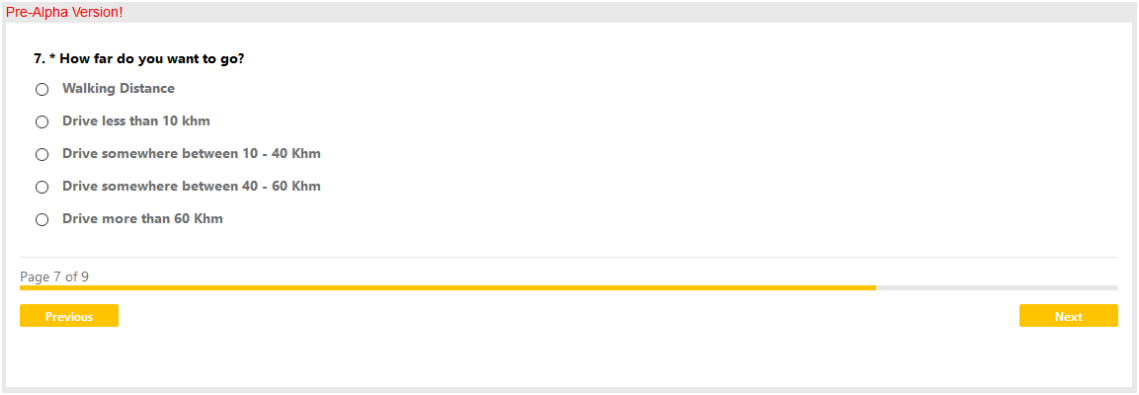

Figure 18: Question number 7 as seen on ExoDecision platform

Question number eight, targets the style of the place the user prefers to be assign to. In addition, we can see clearly in Figure 19 that we tried to be as flexible as we could with the options. We implement seven options thus the user can mark a certain option depending on their needs of style.

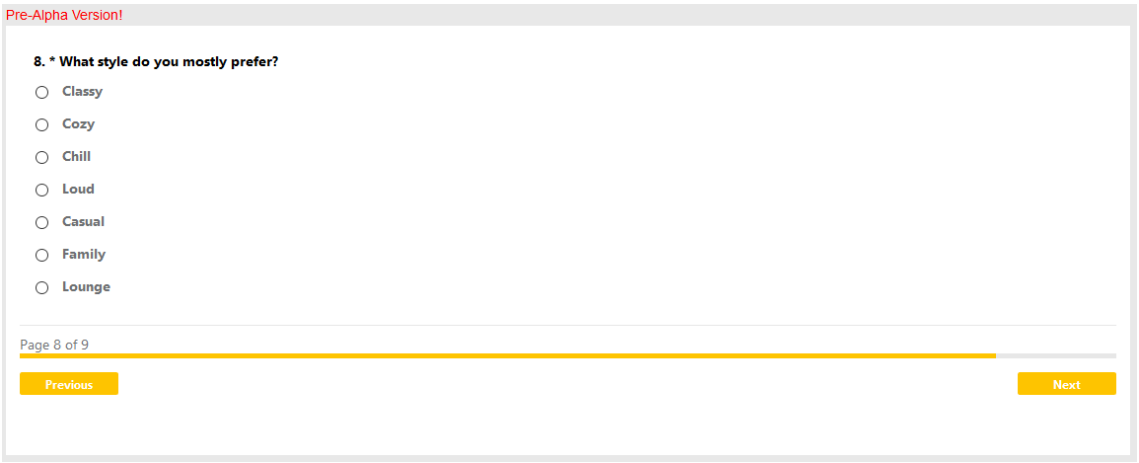

Figure 19: Question number 8 as seen on ExoDecision platform

Finally, question number nine which is the most vital for the Pre-Alpha version of the platform, describes the current sequence of answers applied at that time. The user chooses the proper category, depending on their judgment and on their sequence of answers they provided. Later on, we can use this data for training the machine learning.

Eventually, at some point, question number 9 will be removed. As we can see in Figure 20, there are nine options available to the user and it is up to them where this sequence of options will be implemented to. Afterwards in the Alpha version of the platform, the ninth question will be replaced by the already trained model that W.E.K.A resulted in the Pre-Alpha version. Then, we will recommend to the users a suggested category for their next exodus depending from their sequence of answers. Moreover, to that matter, as the platform arrives to the Alpha version and because of insufficient time, we will not be able to create a friendly graphical user interface giving the suggestion to the user. Users will be able to see only the category label that machine learning suggested to them. For example, users will only see "Reasonably Priced for Coffee" if the sequence of their answers is resulting to that exact category.

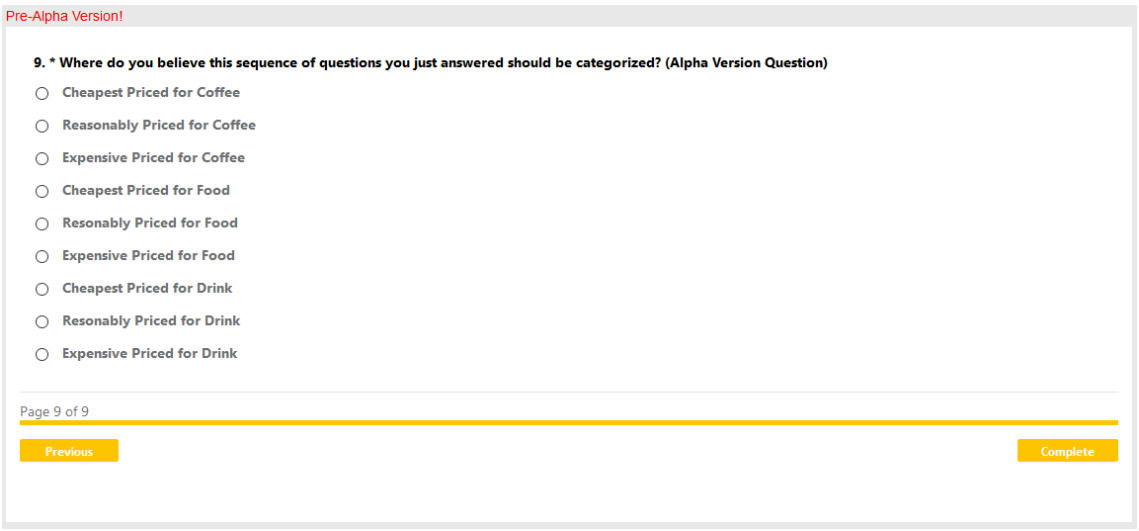

Figure 20: Question number 9 as seen on ExoDecision platform

Therefore, it is worth noting that the data such as "How expensive is a place?" , "What music does it plays?", "Where is it exactly located?" and "What type is it?", have been fetched using the popular platform Foursquare which is a crowd sourcing platform depending on the reviews, ranking and comments of a certain place by users. Foursquare has been in the World Wide Web since November 2011. Their main purpose is to provide to the user's experiences by other users of a certain place that have been there. It is a world-wide platform and it is being used as mentioned in their website by more than 50 million people. (Foursquare, 2018)

7.4 Machine Learning and Performance Evaluation

After acquiring some representative entries of completed questionnaires, we had to proceed to the next stage where the dataset (results) will be imported to the program W.E.K.A for creating a model through machine learning.

Each time a user was completing the questionnaire, their entries had been stored in-to the MongoDB to a collection as a JSON entry. As we can see clearly in Figure 21, the platform was saving the "_id" of the questionnaire which it was created by different bytes values such as seconds since the Unix epoch, machine identifier, process id and a counter (MongoDB, 2018). The "user" id was created the same way as "_id" we mention before, "answers" is the section where the answers were recorded as a comma separated values for easier export to csv or JSON, and finally the "created" value is when (Date/Time) the questionnaire was submitted to the MongoDB.

We exported the stored data through the mongoexport for creating a CSV which will then be inputted to WEKA for further processing.

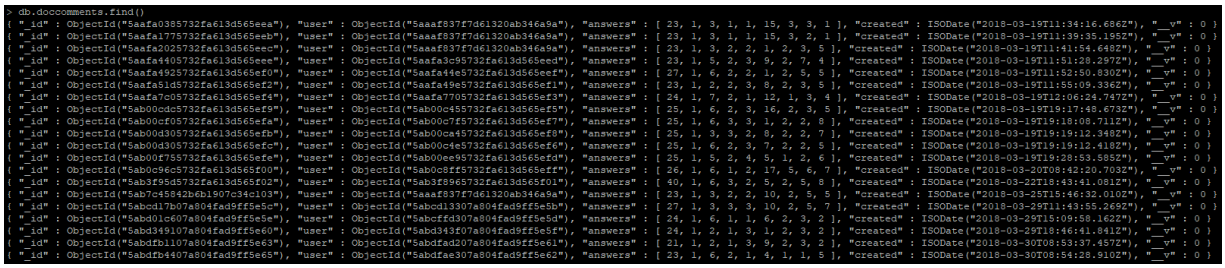

Figure 21: MongoDB data shown for DocComments collection

After the export of the 88 entries of the data set to csv file, we import it to W.E.K.A for creating a model, through machine learning, capable of classifying entries to one of the exodus category options available. We used the Classify function called "MultilayerPerceptron" because, MultilayerPerceptron can identify a weight (Number) in each connection. For example, in Figure 23, we can see that the option number 5 has a total weight of 28. This means that the specific option has been selected 28 times. It is worth mentioning that we have tried other classifications, but they did not function as Multilayer Perceptron. We also applied a filter that transforms Numeric to Nominal since our data is numeric we have to apply this filter in order to convert the types and so on the function can properly work with the dataset.

To illustrate the process of W.E.K.A produce, we can further understand on Figure 22, how the data are being showed. W.E.K.A is showing that after the test it managed to find 42 Correctly Classified Instances with a percentage of 47.7273%. Furthermore, it found also 46 Incorrectly Classified Instances with a percentage of 52.2727%.

```
Time taken to build model: 1.9 seconds

=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances          42               47.7273 %
Incorrectly Classified Instances        46               52.2727 %
Kappa statistic                          0.3402
Mean absolute error                      0.1159
Root mean squared error                  0.2984
Relative absolute error                 64.4474 %
Root relative squared error             99.9149 %
Total Number of Instances               88

=== Detailed Accuracy By Class ===

                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
                 0.200    0.115    0.182      0.200    0.190      0.081    0.555     0.224     1
                 0.591    0.121    0.619      0.591    0.605      0.477    0.835     0.613     2
                 0.000    0.000    ?          0.000    ?          ?        0.483     0.022     3
                 0.000    0.074    0.000      0.000    0.000      -0.080   0.707     0.145     4
                 0.643    0.200    0.600      0.643    0.621      0.435    0.806     0.717     5
                 0.000    0.060    0.000      0.000    0.000      -0.054   0.500     0.060     6
                 0.200    0.024    0.333      0.200    0.250      0.224    0.708     0.326     7
                 0.800    0.051    0.667      0.800    0.727      0.692    0.977     0.834     8
                 0.000    0.000    ?          0.000    ?          ?        0.218     0.014     9
Weighted Avg.    0.477    0.123    ?          0.477    ?          ?        0.766     0.535

=== Confusion Matrix ===

  a  b  c  d  e  f  g  h  i   <-- classified as
  2  4  0  0  1  1  0  2  0 |  a = 1
  3 13  0  1  3  1  1  0  0 |  b = 2
  0  1  0  0  0  0  0  0  0 |  c = 3
  2  0  0  0  4  1  0  0  0 |  d = 4
  2  1  0  5 18  2  0  0  0 |  e = 5
  1  0  0  0  3  0  0  0  0 |  f = 6
  0  2  0  0  0  0  1  2  0 |  g = 7
  1  0  0  0  0  0  1  8  0 |  h = 8
  0  0  0  0  1  0  0  0  0 |  i = 9
```

Figure 22: W.E.K.A 1st results of the data set

As we can understand from these results, machine learning is not performing well. There is a higher probability to recommend something which is not correct rather than something that is correct. This happens because, out dataset is imbalanced, since in question number nine, there are answers which have less selections by users rather than others. With further information, as we can see in Figure 23 which is the visualization of question number nine, option number one has been selected 10 times out of 88,

option number two has been selected by 22 users out of 88, option number three was selected only by 1 user out of 88, option number four has been selected by 7 users, option number five has been selected the most with 28 users selecting that option, 4 users selected option number six, 5 users have selected option number seven, 10 users selected option number eight and again only 1 user selected the option number 9. In other words, our dataset was very small resulting to this problem.

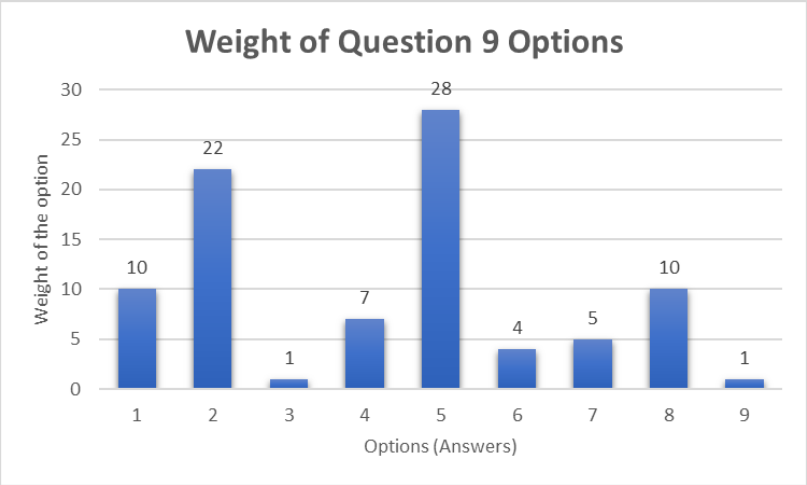

Figure 23: Visualization of question 9 option weights

As we can see, we have scattered results in question number nine. This does not help the machine learning to determine what to recommend to the user. We believe that more balanced answers would have given us more accurate results.

Further steps we need to make in order to resolve this issue of low correctly classification percentage, is to minimize the dataset as much as possible, since we do not have more time to redirect the website to more random users.

Additionally, we removed from the question number nine the entries with the fewer entries. To be more precise, we removed seven out of nine options to minimize the data set. We kept only the options number two and number five since they had the most results in an option.

Moreover, when we retried to import it to W.E.K.A, we found out that we have questions also having less entries at some options. For example, as we can see in the visualization of the first question in Figure 24, we can understand that there are some options that weight more than others.

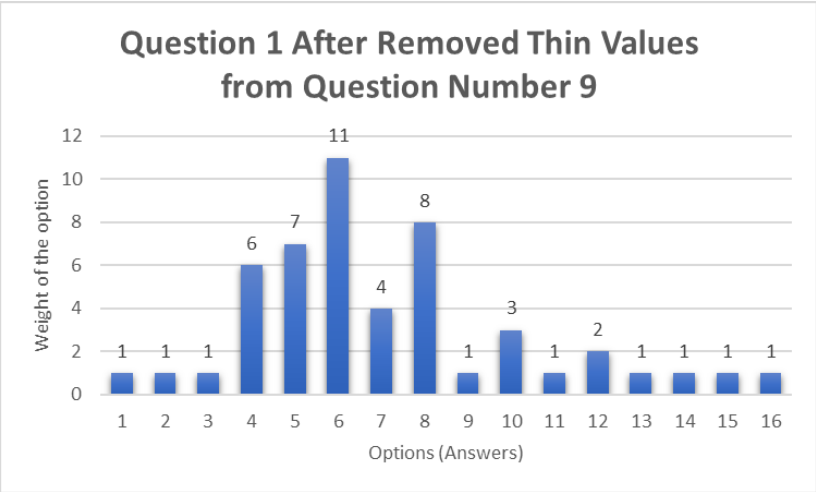

Figure 24: Question 1 after removed thin values from question number 9

In Figure 25, we can see the question number two in which, as we mentioned before, by default is the option number one which is Limassol.

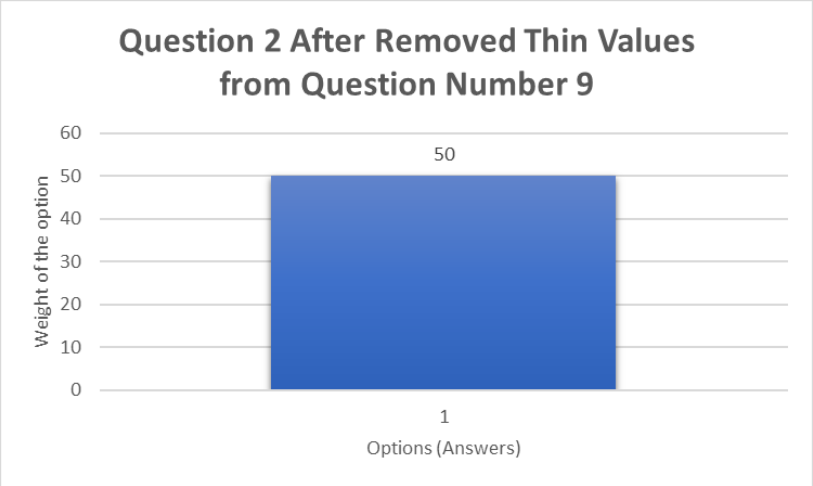

Figure 25: Question 2 after removed thin values from question number 9

In Figure 26, question number three, we can understand that option number one, option number three and option number four, weight less than option two and option five.

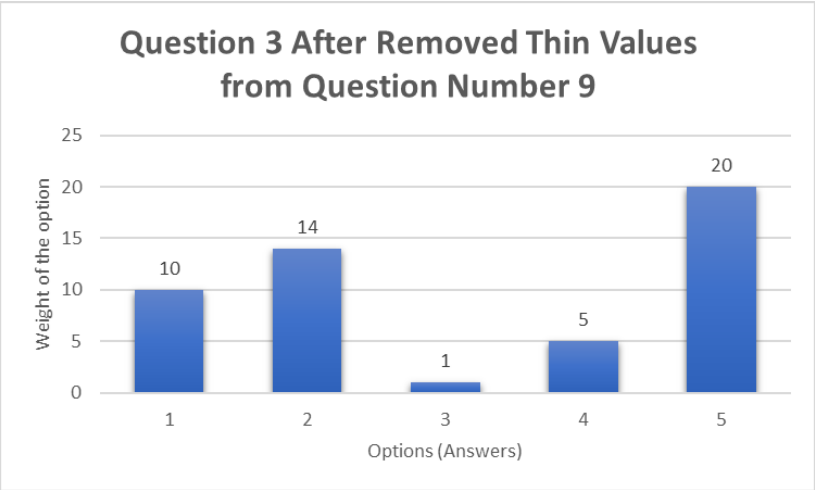

Figure 26: Question 3 after removed thin values from question number 9

In Figure 27, question number four, we can understand significantly that option number three is way much less selected rather than option one and two.

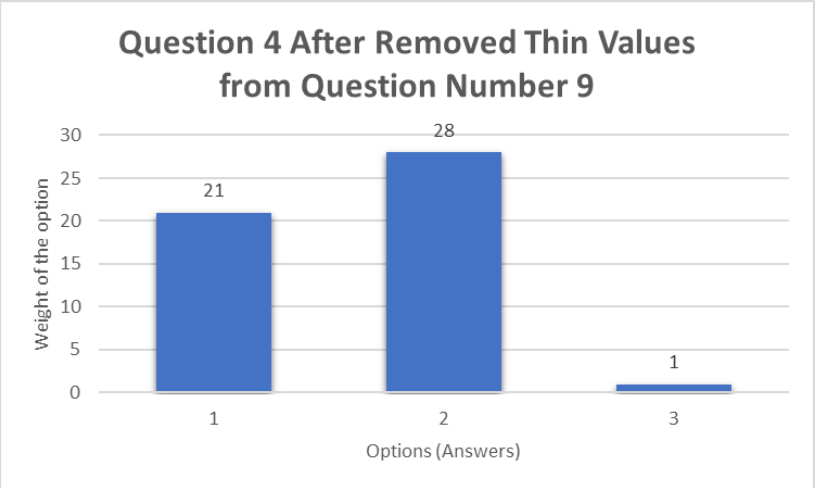

Figure 27: Question 4 after removed thin values from question number 9

Question number five, as we visualized it in Figure 28, option number one, two, four and five are weight less rather than option number three.

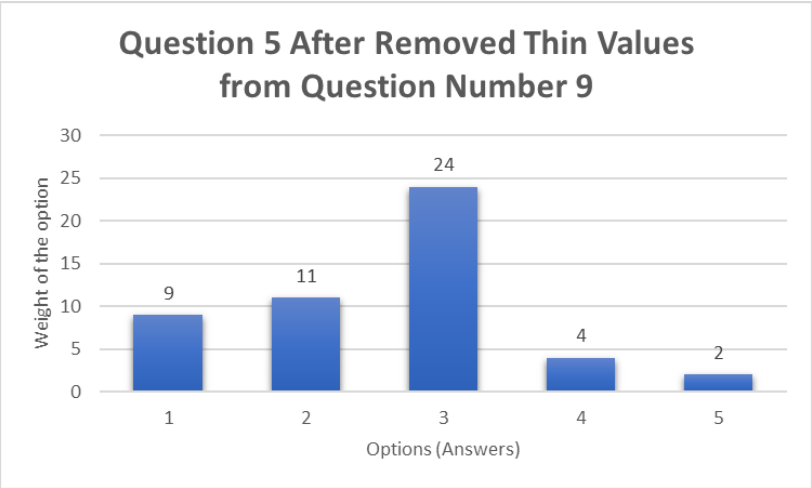

Figure 28: Question 5 after removed thin values from question number 9

Question number six, has many more options compared to other questions. Options one, nine and ten are the most popular choices among users as we can see in Figure 29.

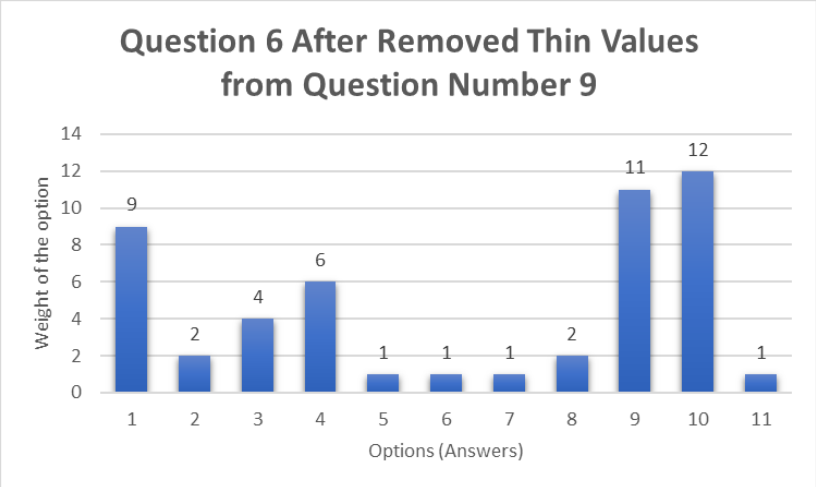

Figure 29: Question 6 after removed thin values from question number 9

Question number seven, has only four options. Only one option had the most weight and it was number two as we can see in figure 30.

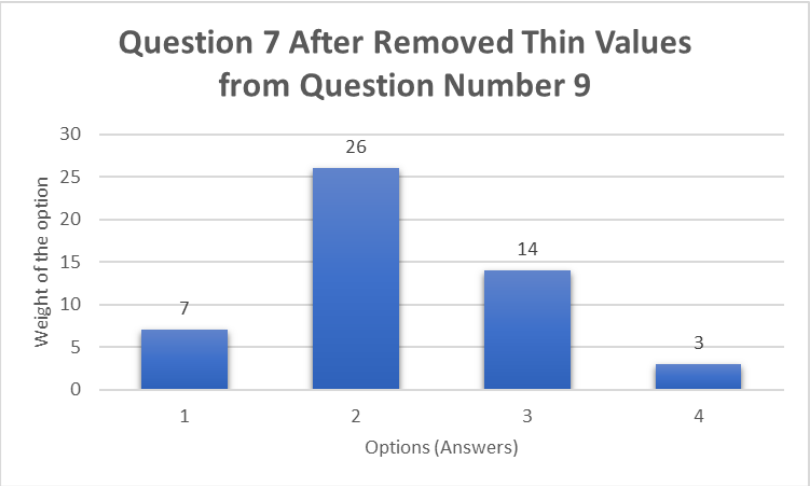

Figure 30: Question 7 after removed thin values from question number 9

Moreover, question number eight had three options with approximately equal weight. As we can see in Figure 31, the other three options are way below than the rest options. Finally, question number nine, is balanced after the removed options one, three, four, six, seven, eight and nine.

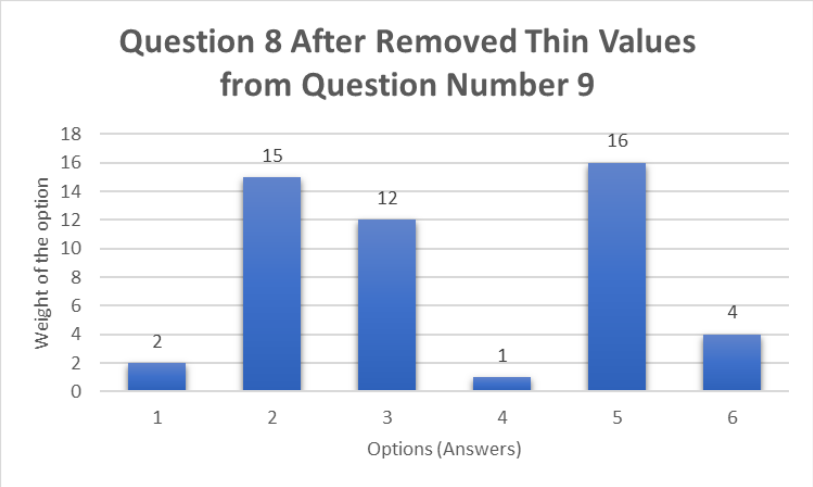

Figure 31: Question 8 after removed thin values from question number 9

Depending from the stats we received, we must remove the options that are not high enough on entries, because machine learning needs to have a lot of data in several options in order to find correct classified instances at the end of the testing. In addition, this is a proof of concept that it can actually work regarding the very small dataset we gained. In order to fully function a machine learning is to have more than 1000 entries of dataset.

Furthermore, we managed to delete all the above options that did not had high weight, and we re-run the test with the same function but this time with the total of 13 entries. As we can see in Figure 32, we've managed to hit a decent percentage of Correctly Classified Instances which was 69.2308%. Additionally, the Incorrectly Classified Instances was 4 out of 13, hitting the 30.7692%.

```
Time taken to build model: 0.03 seconds

=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances           9               69.2308 %
Incorrectly Classified Instances         4               30.7692 %
Kappa statistic                          0.2778
Mean absolute error                      0.2804
Root mean squared error                  0.4515
Relative absolute error                 62.2273 %
Root relative squared error             94.0362 %
Total Number of Instances               13

=== Detailed Accuracy By Class ===

               TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
               0.500    0.222    0.500      0.500    0.500      0.278    0.750     0.667     2
               0.778    0.500    0.778      0.778    0.778      0.278    0.750     0.879     5
Weighted Avg.  0.692    0.415    0.692      0.692    0.692      0.278    0.750     0.814

=== Confusion Matrix ===

 a b   <-- classified as
 2 2 | a = 2
 2 7 | b = 5
```

Figure 32: Removed options from questions that did not had high weight values

CONCLUSIONS

In this section we are going to discuss the results in relation to our research question. Our research question as we mentioned before was, “Is it possible to create a platform where machine learning is implemented, in order to give an accurate recommendation to the user?”. We have managed to create a user-friendly, easy to use interactive and fast platform by using M.E.A.N stack with the compatibility corresponding to smartphones screens. We implemented the questionnaire to the platform and managed to have 78 user accounts sign up to our platform. Furthermore, we managed to have 88 completed questionnaires which at the end did not help us a lot with the research because, as we mentioned before, machine learning needs considerably more data in order to improve its performance so that it can accurately provide recommendations to the users. We managed to create a simple model with only 13 out of 88 which resulted to a more accurate recommendation. Furthermore, we did not have enough time to create the design prospective of the results impending to the screen of the user, but we created a more simplistic way of reading the results as a plain text. In other words, the platform is complete without any serious problems regarding the programming part. Moreover, the platform can be expanded to a more bigger aspect of problem solving in the community.

The focus of the study was to create a platform on which users could connect from any place any time. The platform should help people decide more easily their next exodus and by achieving that, users had to go through 8 simple questions in order to help the machine learning to understand their exact mood. At the end of the questionnaire, users will actually get a category of recommendation for a place such as a café, restaurant or bar which suits with their mood.

It is worth to mentioning that, some restrictions and limitations of the study we’ve been through.

Firstly, we have managed to complete the platform within 3 months, design wise and programming wise, but the problem was that we were left with a limited amount of time for receiving a satisfying number of answered questionnaires. As we mentioned before, machine learning needs a huge dataset in order to work and actually learn from the data. If we had more time to focus on the users, it will be better to have more than a thousand

results, to make sure machine learning can understand and suggest correctly a recommendation.

Moreover, it will be positive if we created also the platform in Greek language, because the base of the platform it is in Cyprus and the official and mostly spoken language there is Greek.

Furthermore, it is worth to mention that, it will be better to have an easier domain name in order to be more user-friendly and easier to remember. A good example will be www.exodecision.com rather than vaa.cut.ac.cy:8081.

Finally, we will suggest some ideas for further research and future work. Primarily, as we mentioned in the methodology chapter, we could arrange a feedback at the end of the suggestion, asking the users if the recommendation was an accurate after their visit to the suggested place in order to get a result if the accuracy of the machine learning model was accurate.

Next, as we mentioned in the ethical issues chapter, will be essential for future work to make a filtering of the options available in the questionnaire. This will help for the scenario where under 18 users will not be able to see options that are relevant to only adult drinks places such as bars and clubs.

REFERENCES

Bobadilla, J., Ortega, F., Hernando, A., & Gutiérrez, A. (2013). Recommender systems survey. *Knowledge-Based Systems*, *46*, 109–132. https://doi.org/10.1016/j.knosys.2013.03.012

Brozovsky, L., & Petricek, V. (2007). Recommender System for Online Dating Service. *Unpublished Manuscript*, 1–12. Retrieved from http://arxiv.org/abs/cs/0703042

Chaniotis, I. K., Kyriakou, K.-I. D., & Tselikas, N. D. (2015). Is Node.js a viable option for building modern web applications? A performance evaluation study. *Computing*, *97*(10), 1023–1044. https://doi.org/10.1007/s00607-014-0394-9

Express.js. (2018). Express.js. Retrieved April 28, 2018, from https://expressjs.com/

Foursquare. (2018). Foursquare About Us. Retrieved April 28, 2018, from https://foursquare.com/about

Gyorodi, C., Gyorodi, R., Pecherle, G., & Olah, A. (2015). A comparative study: MongoDB vs. MySQL. *2015 13th International Conference on Engineering of Modern Electric Systems, EMES 2015*, 0–5. https://doi.org/10.1109/EMES.2015.7158433

Internet Live Stats. (2017). Number of Internet Users (2016) - Internet Live Stats. Retrieved from http://www.internetlivestats.com/internet-users/

Jain, N., Mangal, P., & Mehta, D. (2015). AngularJS: A Modern MVC Framework in JavaScript. *Journal of Global Research in Computer Science*, *5*(12), 17–23. Retrieved from http://www.jgrcs.info/index.php/jgrcs/article/view/952

Kadam, A., Chaudhari, H., Patil, C., & Chavhan, S. (2017). Web Application Development by Using MEAN Stack. *International Journal of Science Technology Management and Research*, *2*(4), 15–18.

Katakis, I., Tsapatsoulis, N., Triga, V., Tziouvas, C., & Mendez, F. (2012). Clustering online poll data: Towards a voting assistance system. *Proceedings - 2012 7th International Workshop on Semantic and Social Media Adaptation and Personalization, SMAP 2012*, 54–59. https://doi.org/10.1109/SMAP.2012.19

MEANJS. (2017). MEANJS. Retrieved April 19, 2018, from

http://meanjs.org/docs.html

MongoDB. (2018). ObjectId - MongoDB Manual. Retrieved June 4, 2018, from https://docs.mongodb.com/manual/reference/method/ObjectId/

OWASP. (2017). OWASP Implement Proper Password Strength Controls. Retrieved April 29, 2018, from https://www.owasp.org/index.php/Authentication_Cheat_Sheet#Implement_Proper_Password_Strength_Controls

SourceForge. (2018). Weka - Download Statistics. Retrieved April 12, 2018, from https://sourceforge.net/projects/weka/files/stats/timeline?dates=2000-04-26 to 2018-04-12&period=daily

Sparks, A. (2017). 5 companies that are dominating with machine-learning. Retrieved April 8, 2018, from https://www.domo.com/blog/5-companies-dominating-machine-learning/

Wang, S., & Summers, R. M. (2012). Machine learning and radiology. *Medical Image Analysis*, *16*(5), 933–951. https://doi.org/10.1016/j.media.2012.02.005

Wiederhold, G., & McCarthy, J. (1992). Arthur Samuel: Pioneer in Machine Learning. *IBM Journal of Research and Development*, *36*(3), 329–331. https://doi.org/10.1147/rd.363.0329

APPENDIX I

ExoDecision Official Link

http://vaa.cut.ac.cy:8081

APPENDIX II

ExoDecision Platform Full Programming Code at Communication and Internet Studies Servers

http://cis.cut.ac.cy/~as.stavrinos/ExoDecision_full-code/