UNIVERSITY OF WESTMINSTER

INFORMATICS INSTITUTE OF TECHNOLOGY
In collaboration with
UNIVERSITY OF WESTMINSTER

5COSC009C - Software Development Group Project
Software Requirement Specifications

Themis

Cave Coders

Name	**IIT ID**	**UoW ID**
Rashmi Senanayake	2018426	w1742120
Aaron Peries	2018414	w1736298
Pasindu Ekanayaka	2018405	w1742110
Ravindu Ratnayake	2018440	w1742310
Shiom Perera	2018441	w1742311
Tony Jesuthasan	2018596	w1743058

Table of Contents

Table of Contents

Table of Contents

Table of Figures

Table of Figures

8. Appendix

List of Tables

List of Tables

1.Introduction

1.1. Introduction to the Problem

1.1.1 Problem Boundary

The project at hand is limited to the legislature of the Democratic Socialist Republic of Sri Lanka. Furthermore, the judgements used for the project will be those adjudicated by the Supreme Court of Sri Lanka and focused on federal and civil cases.

The verdict given by a judge in court can be used in a current court case. This is known as judicial precedent (Landes and Posner, 2020). A lawyer faces problems when trying to retrieve the verdicts of past cases due to lack of time,human nature and human errors.

Lawyers can use these previous verdicts to lessen the sentence given by the judge or even have no sentence given at all. Nobody wishes to be wrongly convicted and this system would help defendants prove themselves worthy of a lesser sentence or none at all.

1.1.2 Example Scenario of the Problem

Let us imagine an example scenario where a particular man Mr A was riding a motorbike along the Galle road in Wellawatte and was very careless and negligent while riding it. This particular person however gets involved in an accident where he collided with a bus resulting in fatal injuries. Within a few minutes after the accident the people around managed to remove this particular person, Mr A from the scene and admit him to the nearest hospital. At the time of the crash a young woman, Mrs B, at that time 7 months pregnant, was around 50ft from the scene of the crash and heard it and was shocked. Then she arrived to the scene of the accident and saw a pool of blood and was further shocked. Later she gives birth to a stillborn child, and claims that she had suffered psychiatric harm and sustained loss due to this particular man Mr A.

Now this particular person Mr A has approached the lawyer Mr.C, seeking legal help and advice. However Mr.A clearly holds responsibility for the sustained loss during the crash due to his negligence and carelessness. On top of that the young lady Mrs.B claims that a duty of care had been owned to her by Mr A .

How can the lawyer Mr.C legally help Mr.A who has approached him for help? There is a very clear area to prove that Mr.A holds responsibility for the damage. This is where our application will help both the lawyer (Mr.C) and the client (Mr.A). The only way the lawyer can bring some sort of reduced punishment to Mr.A is by finding the right verdicts and proving that Mr.A deserves that judgement too. The lawyer Mr.C can enter the keywords of the case he is dealing with and our application will provide the lawyer the most relevant and similar cases along with the verdicts given thus immensely benefiting the lawyer.

For example for the above scenario the two cases and verdicts that will be provided by our system to the lawyer will be as follows.

(1)Bourhill v. Young , Scottish delict case , 1943.

Mr Young who had been negligently riding a motorcycle along a road got involved in a collision with a car, fatally injuring him. At the time of the crash, Mrs Bourhill was about to leave a tram which she had been riding, around 50ft from the scene of the accident and heard the crash. Following the removal of Mr Young's body from the road , she approached the scene of the accident, seeing the blood remaining from the crash. Mrs Bourhill, at the time eight months pregnant, later gives birth to a still born child, and claimed she had suffered nervous shock, stress and sustained loss due to Mr Young.

Judgement by Lord Russell at the House of Lords :

"Can it be said that John Young could reasonably have anticipated that a person, situated as was the appellant, would be affected by his proceeding towards Colinton at the speed at which he was travelling? I think not. His road was clear of pedestrians. The appellant was not within his vision, but was standing behind the solid barrier of the tramcar. His speed in no way endangered her. In these circumstances I am unable to see how he could reasonably anticipate that, if he came into

collision with a vehicle coming across the tramcar into Glenlockhart Road, the resultant noise would cause physical injury by shock to a person standing behind the tramcar. In my opinion, he owed no duty to the appellant, and was, therefore, not guilty of any negligence in relation to her."

(2) Case of Donoghue v. Stevenson

Lord Atkin formulated the 'Neighbor Principle' and the concept of duty of care was elaborated.

The three elements of the principle :

1. Proximity
2. Foreseeability
3. Standards of care.

1.1.3 Attempted Solutions

On a local scale, there are a few online websites which provide related cases when searched for, namely "www.lawlanka.com" and "www.lawnet.gov.lk". However the searches on these websites can only be conducted based on a few words which are entered by the user which make the results very vague. There are no systems to predict the certainty of a case's failure or success in Sri Lanka and also provide the most likely verdict to be delivered.

1.2. Aim

The main aim of this project is to design and develop an automated system that would do the following:

It hopes to help a lawyer find identical case judgements to the one he/she is handling by entering keywords and filtering the search according to their needs.

Predicting the probability of the success of the case by taking into consideration the legislation of Sri Lanka and the success or failure of similar cases which have taken place.

Delivering the most likely verdict that would be given once again based on the legislature of Sri Lankan and the materials of a similar case.

1.3. Scope

1.3.1 In-Scope

- Lawyers have to enter through a login page. They can sign in to the system upon first visit to the website by providing Email address, law license details and their own password.
- Creating a user friendly and attractive website.
- Creating a website that user(lawyers) can find previous cases which are similar to their present or future cases.
- Search for similar cases by keywords or by uploading the case file.
- Filter by year, statutes, jurisdiction.
- Case files should be in english.
- System supports english language.
- Show lawyers the most-likely verdict if case files are uploaded
- Provide lawyers with the probability of winning the case based on the past case files.

1.3.2 Out of Scope

- Implementing the system that compatible with languages other than English.
- Implementing a system that is compatible with other national legal systems.
- Provide the user with the mean value of the case duration.
- Download case file

1.4. Objectives

1.4.1 Research Objectives

This project is mainly focused on helping the lawyer in two different manners one is in, when a case is taken by the lawyer, the system finds similar cases with the judgement to help the lawyer understand how to approach the case. Most lawyers can't spend time on finding similar old cases due to the time limit and it's unrealistic nature, due to many cases are present of different sections. But with our system using NLP we find similar cases for the lawyer to use making their job easier.

And the other way of helping the lawyer is, giving him the probability of him winning the case and the most possible verdict which would be delivered based on the legislature of Sri Lanka and similar cases which have been closed earlier, thus deeming to be helpful for the lawyer whether to take the case or not. Furthermore, using these insights the lawyers will be able to plan their case strategy accordingly.

1.4.2 Academic Objectives

1. Learning how to process and analyse unstructured data.
2. Learning how to categorize data into required rows and columns.
3. Learning how to identify patterns and sequences using unstructured data and manipulate them to get the required output.
4. Learning to use new technologies to design the front end according to the required specifications.
5. Learning to use new technologies to design front end and manipulate the front end around the backend to make the site fully functional .

6. Learning how to use frameworks and server side technologies to produce the expected outcome using the structured data.
7. Learning how to use databases and manipulate the data in the databases with the application ui created.
8. Learning how to use Natural language processing.
9. Learning how to write various API's.
10. Working with Machine Learning.

1.4.3 Operational Objectives

When considering the operational objectives, due to our project outcome is based on data we should collect lots of data and categorize it into tables and rows in order to produce expected outcomes in the future.

1.5. Requirements

1.5.1 Rich Picture Diagram

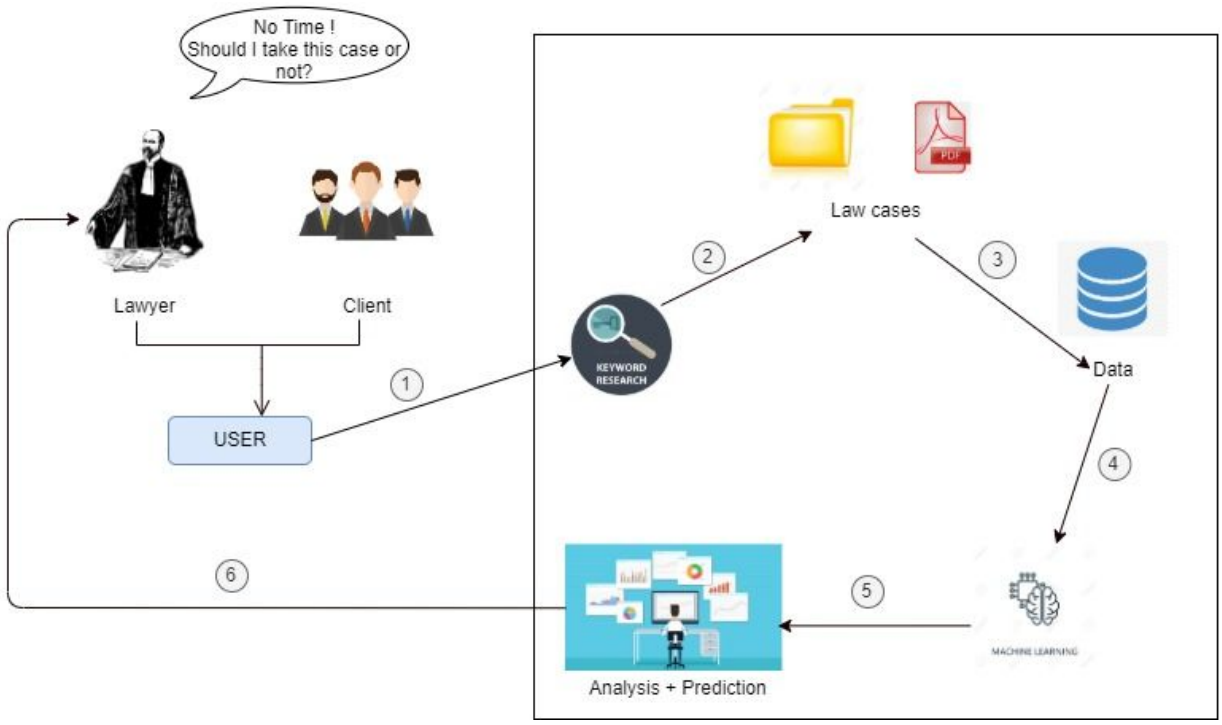


Figure 1.1. Rich Picture Diagram

The above rich picture explains the process of the themis application which has the goal to show similar cases, most likely verdict, winning percentage of a given case.. The internal environment which includes the functions and the features of the application is surrounded by a boundary. The flow of the processes is shown by arrows in a sequence shown by numbers. The callout above the lawyer says the reasons for lawyers to use this application.

1.5.2 Features of the prototype

The prototype will be able to do the following:

1. Display relevant results when a user searches for similar cases using keywords, filters and the case file.
2. Calculate and display the probability of a case being won or lost.
3. Analyze and display the most likely verdict that would be delivered.

1.5.3 Software Requirements

Operating System	● Windows 7 (32-bit) or higher ● Mac OS X 10.8 or higher
Web Browsers	● Google Chrome ● Microsoft Edge ● Mozilla Firefox ● Safari
Browser Compatibility	● Javascript must be enabled. ● Cookies must be enabled.

Table 1.1 - Software Requirements

1.5.4 Hardware Requirements

RAM	512MB or higher
Processor	1.5GHz or higher clock speed.
Devices	Keyboard and a mouse or any other compatible pointing device.
Internet Connection	Broadband (high-speed) Internet connection with a speed of 1 Mbps or higher.

Table 1.2 - Hardware Requirements

1.5.5 Technical Stack

Angular	To develop the client side part of the system.
PlayFramework/ Spring/ Spring Boot	To develop the server side part of the system
Javascript, HTML, CSS	To create a web application.
Python	Using for machine learning processes.
TensorFlow	To develop Machine Learning Schemas.

Table 1.3 - Technical Stack

2.Literature Review

2.1. Chapter Overview

Court case verdicts are an integral part of society as they are the fruits of justice. It is well known that a verdict that has been delivered before in a higher court can be used by lawyers in a lower court during case proceedings if the two cases share similar characteristics. This is known as Judicial Precedent. Searching for similar verdicts can be quite difficult as it requires lawyers to read through many cases before deciding if it is suitable for the current case. This is a very time consuming process and lawyers may sometimes miss a similar case due to human error. A software that eradicates this troublesome process would help lawyers have more material during court proceedings and also increase the chances of having the most just verdict delivered. Furthermore, it is quite difficult for a lawyer to provide the probability of a case winning or losing by looking at the case details that are provided to him/her initially during the start of a case. A software system that provides the probability of the case based on the past verdicts delivered and the laws of the nation and the current case details. This allows lawyers to decide if they want to handle the case or not. Knowing the most possible verdict beforehand will be handy for lawyers to decide the path they take during the case proceedings. Thus once again having a software system that allows lawyers to have the most possible verdict based on previous similar verdicts, the law of the nation and the current case details would help them.

This Literature Review discusses the existing work that has been carried out in the domains, technologies and existing systems related to the above software features mentioned. The pros and cons of various advancements and their evaluation.

2.2. Literature Review of Domain

2.2.1 Introduction to Natural Language Processing

Natural Language processing in simplest terms is the ability for computers to understand what we humans use to communicate with each other, Language. Today, Natural Language Processing is being used by us almost everyday without our knowing. Spell checking, spam filters, autocorrect, and voice assistants are basic everyday use cases. Natural Language processing can be broken down into two parts which are Natural Language Understanding and Natural Language Generation (Diksha Khurana, Aditya Koli, Kiran Khatter and Sukhdev Singh, n.d.) . Natural Language Understanding (NLU) like the name suggests focuses on the comprehension of Natural Language. Whereas, Natural Language Generator (NLG) is the process which produces a set of words which have meaning together such as phrases, sentences and paragraphs. Now knowing what Natural Language Processing exactly means and its sub-areas we should now understand how it works.

2.2.2 How does Natural Language Processing work?

In almost all the languages many words have different meanings.These words are called Homonyms. In the English Language for example the word “Right” has not one not two but three different meanings. It can be used to mean that something is correct, the direction or a moral or legal entitlement that someone is entitled to have. With such words in existence the difficulty of teaching the computer to understand increases.(IONOS Digitalguide,2019) Computer linguists rely on the five linguistic fields available which are morphology, syntax, semantics, phonology and pragmatics. These linguistic fields have different methods which are used to derive knowledge. Tree Diagrams and Part-of-Speech (PoS) tagging are two of these methods. These methods of deriving knowledge from these linguistic fields are a very detailed

process which is very complex to explain in a paragraph or two. Instead of this we will see how natural language processing is used in our project later on.

2.2.3 Introduction to Machine Learning

Humans and animals have brains that have different capabilities on their own and to use that it should be trained. Similar to the brain, computers can learn from experience and use that information in the future. In order to do that we use machine learning. Algorithms in machine learning used for teaching computer methods to learn directly from data without relying on manual instructions coded by the user. We use machine learning when there's a hard task involving huge number of data to work with. Computer finance, biology, energy production are some of the fields that use machine learning.

2.2.4 How does Machine Learning work?

Data is the most important component for machine learning without data, we can't teach computers anything and luckily there's plenty of data to use, but dealing with a huge number of data (big data) can be tricky. Supervised learning and unsupervised learning are the two techniques that machine learning use. In supervised learning, it trains according to data that we know what the input and output. In unsupervised learning, there's nothing we know so the model should find hidden patterns and decide the output of its own. The accuracy of the final result will be high if we use more data and questions.(MathWorks, (2017))

2.3. Literature Review of Technologies and algorithms used

2.3.1 Finding of identical case judgements

To find identical cases user should enter keywords(crime,civil,family) and also they can filter by year, status etc. Using these data system should find the identical cases. In order to do that we use supervised machine learning to automatically process text and identify patterns. we should manually label keywords, years of many cases and make a dataset so we can train the system to identify the difference of each keyword, after training the system it should be able to find similarities in words of new cases so that we can find identical case judgements.

Similarity measure is a function which calculates the degree of similarity between a pair of documents or vectors. We don't have the best similarity measure as yet but there are proposed methods which give us quite reliable percentages of performance which even allow us to rank the documents in similarity wise (Arms et al., 2019).
To measure similarity of judgements we can follow 4 different similarity measure strategies which are elaborated below.

(i) All- term cosine similarity
Here one judgement is considered as one entire vector and these judgements are taken into consideration and used to calculate a TF-IDF score. The vector space model is then used to calculate the TF-IDF frequency and get the score (Kumar et al., 2011).

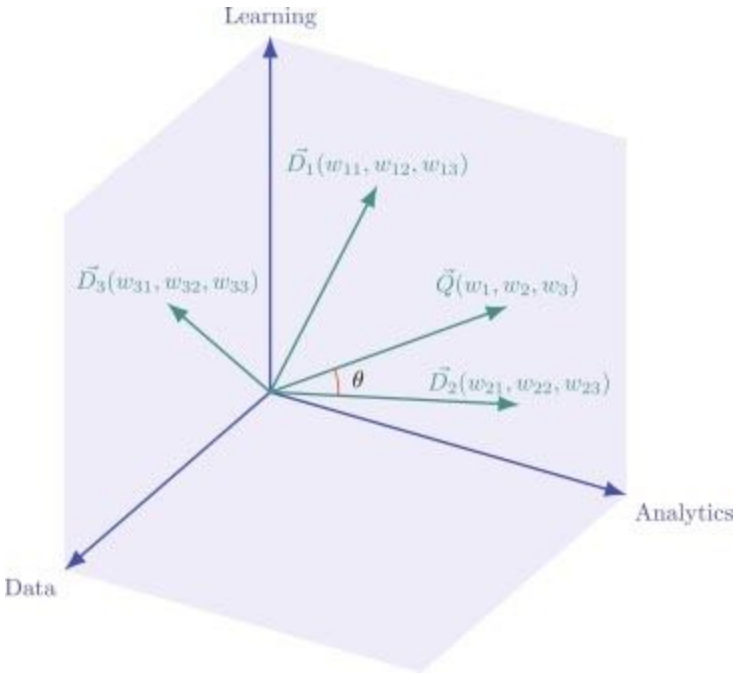


Figure 2.1. Vector Space Model

(Gudivada and Gudivada, 2019)

The above diagram depicts the vector space model that describes the similarity of a document vector D1, D2, D3 with the query Q.

(ii) Legal term cosine similarity

Here we evaluate the legal terms that appear in the legal dictionaries that are used in these judgments. The TF-IDF score will be calculated for these legal terms that we identified using the vector space model and the similarity will be measured.(Kumar et al., 2011)

(iii) Bibliographic Coupling (BC) similarity

In supreme court judgements there are citations referring to other judgments mentioned. These judgments are often known as out-citation. Using regular expressions we can identify out-citations and for each of the judgements a judgement, out-citation pair is prepared. Now we can compare the pairs and select the pairs with the most common out-citations.(Kumar et al., 2011)

(iv) Co-citation (CC) similarity

Here instead of creating a judgement, out-citation we use regular expressions to find in-citations of a judgement. That is selecting other judgement to sections which include the judgement we are working with. Then we create a in-citation,judgement pair and compute the similarity score between the two pairs which also equals to the number of common in-citations.(Kumar et al., 2011)

Out of all these similarity measure strategies legal term cosine similarity and bibliographic coupling similarity have been accepted by domain experts and is more acceptable.

2.3.2 Probability of case success

System decide the probability by considering the previous similar case verdicts. By using the dataset, system find a similar case according the user input keywords and look for the verdict of that previous case. For example, if there are 10 similar cases for the keyword, system check the verdicts of all cases and if 8 out of 10 cases won system will predict 80% success rate for the given case.

2.3.3 Predict most likely court verdict to be delivered

This section focuses on the part of the system which will be able to predict the closest possible verdict that could be given by the court. With such a system non-existing in Sri Lanka, this literature review will act as a benchmark to compare the results of our study to establish such a system in Sri Lanka along with the other findings in this area which have been carried out globally. The most similar document which was found in relation to this part of the system was a research paper by Medvedeva,Vols and Weiling, 2019. The authors use the technologies of natural language processing and machine learning to predict the decisions of the European Court of Human Rights. In order to build our system we would also use the same technologies along with data science. Just like the authors used publicly available data, we too will be relying on publicly accessible data to carry out our development project.

Like every other field, law has its own terminology that helps classify special elements in the field. Using natural language processing, we can identify this pool of keywords used in cases and match them with those used in new cases. (for example, ‘first-degree’) In court, previous cases (also known as the case base) is used to decide the outcome of a new undecided one. NLP can be used to match the most relevant precedents and can draw out a result set of cases containing those cases that will help most with the fate of the currently undecided one. This way, lawyers need not waste time looking for all relevant cases. (also, possible for lawyers to overlook cases that may be relevant due to human errors).

In the proceedings of LOAIT 2010 (“The IV Workshop on Legal Ontologies and Artificial Intelligence Techniques”), Section 1 (Legal Knowledge Extraction), Adam Wyner mentions in his abstract, “*In common law contexts, legal cases are decided with respect to precedents rather than legislation as in civil law contexts. Legal professionals must find, analyse, and reason with and about cases drawn from a set of cases (a case base).*” In this paper, the application of natural language processing tools to cases to produce annotated text is discussed as an initial step. Unlike the system we proceed to create, they don’t intend on delivering the verdicts of similar cases, instead only produce the list of cases that match highly with the current undecided one. (Wyner, A. (2010) 'IV Workshop on Legal Ontologies and Artificial Intelligence Techniques', *Towards Annotating and Extracting Textual Legal Case Elements*, 9)

With the amount of case data available to the general public and law firms, looking for datasets has become less of a burden. When law firms continue to take on cases on behalf of their clients, there is a growing body of case related data in their firms database. This database is used to match keywords of past cases with those of new cases to obtain the verdicts delivered in them, so that they may be used in court to help the cause of the client. (commercial databases like Westlaw and Lexis store cases for future reference). Machine learning algorithms can be applied to find regularities between cases and establish patterns in them. (to completely associate it with a case that matches those patterns in the future).

When dealing with a court case sometimes not all outcomes of court cases can be identified by a program and converted to a binary outcome. Therefore a code can be created to categorise a justice vote into 3 as listed below.

(i) Affirmed

The outcome of the court case is declared affirmed if there are more Affirm votes than than reverse votes.

(ii)Reversed

The outcome of the court case is declared affirmed if there are more Reverse votes than the number of affirmed votes.

(iii)Other

This category holds any vote that cannot be identified as affirmed or reverse.

We can next use the features provided in the datasets to understand the behavior of previous judgements based on different factors such as the term, petitioner, respondent, the reason for granting cert etc. Next we move onto focusing on a result that is derived through interaction of one or more features. One feature we will be looking into is the duration between the argument when it was originally argued and a decision was given. The gathered features then will be processed through a random forest algorithm to give predictions of the outcomes of the court cases.(Katz, Bommarito and Blackman, 2017)

For a prediction like this we can identify an important link inorder to expect results with higher accuracy. A well processed and cleaned dataset is important for the prediction of the most likely verdict to be delivered and based on the predicted result the likely decision will be given as an output.

Data →Prediction→Decision

Out of this mapped relationship the most important relationship is between the data and the prediction. Accurate and well processed data will result in a higher accuracy level of prediction and thus we can rely better on the decisions provided by the model.(Kleinberg et al., 2017)
For the prediction of the most likely verdict to be delivered there are several algorithms that we can follow. However, considering the nature of the problem we work with here we can classify that predicting a verdict given by a court can be a slow judgement. Slow judgments involve a length process of research , reasoning and discussions (Evans et al., 2018).
For our project the algorithm we decided to use is SVD CF (collaborative filtering of singular value decomposition).Collaborative filtering is a methodical process for predicting the rate or the likeliness of user(judge) decision pair based on the history of decisions given by the user and given to the decision (Maher Malaeb, 2016).So a SVD CF algorithm is used to capture the latent relationship between the user and the decision that allows to compute the prediction of the most likely verdict to be delivered by a (judge)user (Vozalis, Markos and Margaritis, 2014). SVD CF has proven to show a higher accuracy rate over other algorithms that can be applied to predict the most likely verdict and surprisingly outperformed neural CF algorithm. Another reason to select SVD CF is because of unavailability of highly classified datasets to work with neural networks algorithms.Our research project is limited to a model that predicts based on the slow judgments that are fed.Therefore the model cannot predict or generalise to new case scenarios that humans haven't answered yet (Evans et al., 2018).

2.4. Literature Review of Existing Systems

For centuries, legal researchers used doctrinal research methods and analysis of past cases to predict or find patterns in the legal sector of a country.(Epstein and Martin 2010). How they carried it out, it was hectic they had to manually collect court cases and their verdicts, read them fully, summarize them and commented to collect their required data only using man power. As

you can see researches have been working on predicting verdicts given in the court. Thus making our project been already existed. Let's observe the existing softwares/projects and see how our project is different from these existing systems, the new additional scopes we are adding and how it affects the final software product.

Researchers in the United States were the first to use machine learning techniques to predict the courts decisions or voting behaviour of judges(Katz 2012; Wongchaisuwat et al. 2017).
Recently, Katz et al. (2017) developed a prediction model that aims to predict whether the US Supreme Court as a whole, affirms or reverses the status quo judgement, and whether each individual Justice of the Supreme Court will vote to affirm or reverse the status quo judgement. Their model achieved an accuracy of 70.2% at the case outcome level and 71.9% at the justice vote level.

Sulea et al. (2017b) used machine learning techniques to analyse case law of the French Court of Cassation. There aim was to predict the law area of a case and the court ruling, which this model achieved the accuracy of above 92%. Using Artificial Intelligence researchers from university college london and the universities of Sheffield and Pennsylvania, predicted the verdicts of the European Court of Human rights.Their system had an accuracy of 79%(Jane Wakefield).Premonition, claims to be the world's largest litigation database, has invented the concept of predicting a lawyers success rate and how long a lawyer will take to each attorney and DISCO claims to deliver faster results for document search and predicts what files needed by the user(Edgar Alan Rayo , 2019).Above are few examples of few similar verdict predicting projects and softwares. As you can all of these only predicts the verdict but our system has few new scopes that enhances all of the projects.

The additional facilities that we provide is that a lawyer can enter keywords of a certain case he/she has to practice so that the system filter and display the relevant similar cases that took place for the past years so that he/she doesn't have to waste time going through manual hardware

to search for past cases. Our system instantaneously displays all the cases so that the lawyer can understand how to approach the case and help his/her client.

Apart from filtering cases and predicting the verdict we also predict the possibility of he/she winning the court case. This functionality will help young lawyers who wants to build a career name by winning many court cases to decide on whether to take the case or pass it to a much more experienced personnel who can do the impossible.

As mentioned above, the existing system all of them are developed for foreign laws whereas our system is applied to sri lanka which is very new as sri lanka has no systems that predict the verdicts or the possibility of success of a certain case. The only system which is similar to our project is the Sri lanka Supreme Court Website where you can search for case law. But this is not very accurate as the search is done using a basic technology whereas in our project we will be using NLP where a search is done using specific keywords and also the website, it doesn't do any predictions making our project more advanced and identical.

2.5. Chapter Summary

This Literature Review exchanged views on the existing domains, technologies and systems to provide a solution to the problem domain. The chapter started off with the review of the existing domains related to the problem at hand, followed by the review of the existing technologies and then the current systems which are related to the problem area.

3.Methodologies

3.1. Chapter Overview

This chapter lists the methods that have been chosen in terms of this project research, the project management and the overall development of the project over time. This section presents the various methods that exist and further discusses why the final methodologies were chosen along with diagrams which show how the project will be implemented.

3.2. Selected Methodologies

As a group of computer science students, our attempt was to solve a problem that affects a particular section in our community with a solution and then to prove that the solution is practically possible. In order to achieve this the project was divided into two phrases which are respectively research and implementation/software development. The process of that each phrase went according to a selected methodology considering the project's time-frame, budget and available resources. The following will discuss those methodologies and explain how those were cooperative to enhance the project's quality.

3.2.1 Research Methodology

Philosophy	There are always two main reasons to conduct a research which are filling knowledge gaps and problem solving. So, in this case we have followed positivism by gathering measurable and comparative data using methods such as questionnaires in order to seek the realistic nature of the solution that we discover by being independent from researchers' perspectives.
Approach	In terms of the research approach, a deductive approach was chosen. According to this approach, we started by developing theories to explain why this problem has occurred and then eventually collected data that confirms it can be solved.

Strategy	Strategy-wise, the best option for this dissertation was survey. Surveys help to find answers on different aspects, and it was useful to gather a large amount of data at a low cost in a short amount of time. In the meantime, interviews were also held to find flaws in the work done.
Choice	From mono, mixed and multi methods, mixed method was chosen as quantitative research was held to find out a practical solution and also a qualitative research was held to understand the nature of the given problem. Eventually this help to provide comprehensive evidence on given event.
Time Horizon	There are two types of time horizons namely Longitudinal and Cross-sectional and cross-sectional was chosen for this project since the implementation/software development will be done after the research.

Table 3.1 - Research Methodologies

3.2.2 Software Development Methodology

When creating any sort of software, it is highly important to plan and manage your work according to an accurate structure. This structure is called as software development methodology. There are several software development methodologies such as waterfall, scrum, agile, spiral, prototype, etc. Among those, Agile Methodology was chosen as the most suitable for this project.

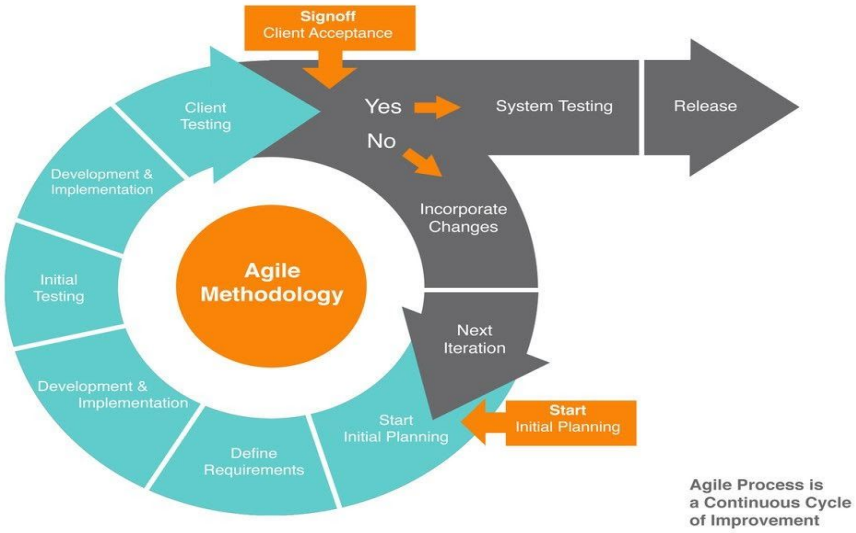


Figure 3.1. Agile Methodology

Agile methodology promotes continuous iteration through design, development and testing phrases. This enables to make changes and fix errors in the middle of the project. So, as a group of students considering the high chance of occurring errors and the changing nature of the requirements, agile became the best choice.

3.3. Work Plan

3.3.1 Activity Schedule

Task	Start Date	End Date	Days
Project Initiation	24/09/2019	05/11/2019	42
Literature Review	05/11/2019	19/11/2019	14
Requirement Engineering	02/12/2019	10/01/2020	38
System Design	13/12/2019	13/01/2020	30

Table 3.2 - Activity Schedule

3.3.2 Work Breakdown Structure

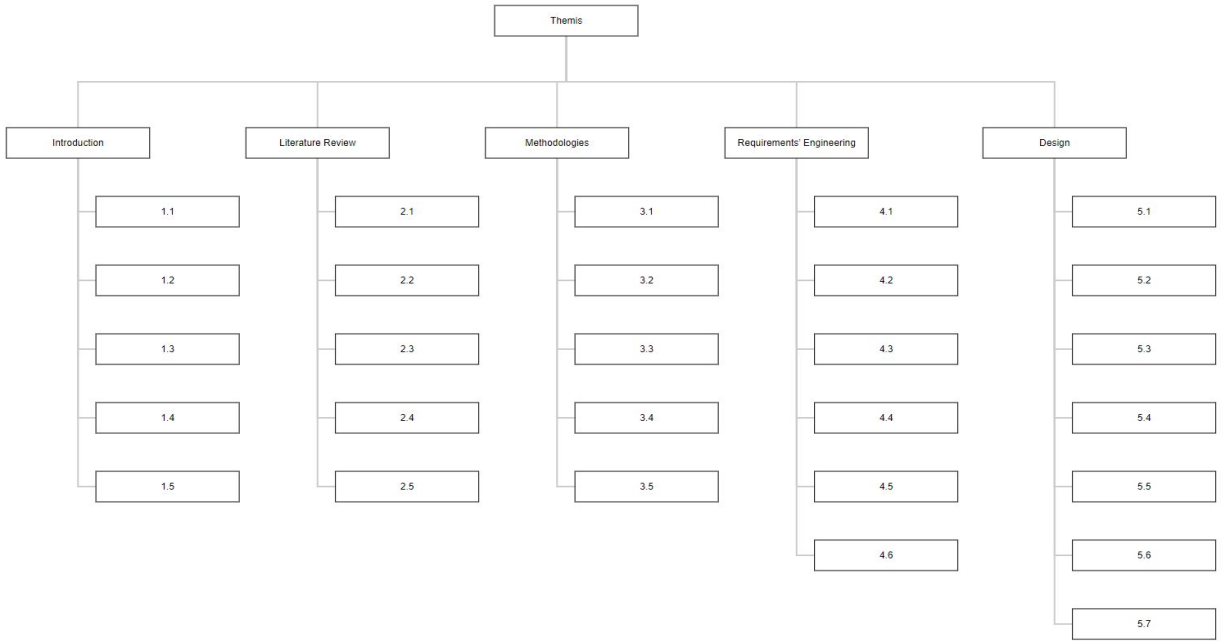


Figure 3.2. Agile Methodology

Please refer to Appendix C for clear diagram.

3.3.3 Deliverables

Deliverable	**Date**
Project Initiation Document	4/11/2019
Literature Review Document	19/11/2019
Software Requirement Specification	27/01/2020
System Design Document	27/01/2020
Prototype	20/04/2020
Thesis	17/04/2020
Review Research Paper	-
Manuscript Paper	-
Public Package / Library	-

Table 3.3 - Deliverables

3.5. Deviations and Risk Mitigation

Any project which has a high time span is subjected have many risks, especially when a project is done for the first time in your career. To overcome and minimize this we should identify and analyze the possible risks that the certain project can have and provide mitigations to override these risks. Below is a table consisting of the risks identified and their relative mitigations.

Risk	Level	Mitigation
Unavailability of Data Sets. for our project the most important component is data science and it requires data sets. If data sets are not available or have no license to access them we cannot proceed furthermore with our project.	High	We should make sure that there is open data sets available and if there is any privacy policy, to get the necessary permission to access it. If data is available in chunks we should summarize them according rows and columns which satisfies our requirement.
Less Knowledge on the Main Component.Unable to identify patterns from our unstructured data we gathered due to poor data analytic skills.	High	Giving more importance and attention to the main component and working on it from the beginning so that there is more time to work on the main component rather than waiting until the end.
Less knowledge on what technologies to use and from where to start. Even after finding the data and also the pattern we should use this to produce our final outcome. For this we should use certain technologies and methods. Less clue on where to start and what to use puts you in great danger	Medium	Planning the start from the very beginning and researching more about the technologies which can be used to satisfy our project and learn them if your not friendly with them.

Passing Due Dates. As this project is our first and the main component is a new varse subject more time will consume due to, us to learn this and work on it at the same time.	Medium	Managing the time properly and breaking the Work equally among the team members and working as a team together helping each other will help us to complete the project on time.

Table 3.4 - Deviations and Risk Mitigation

3.6. Chapter Summary

This chapter took off with the discussion of the selected methodologies for this project, in both areas of research and project management. It went on to discuss the work-plan for the project which included the work breakdown structure, activity schedule, Gantt chart and the deliverables. The deviations and risk mitigations were discussed afterwards. The following methodologies will be followed during the implementation of this system.

4. Requirements' Engineering

4.1. Chapter Overview

This chapter focuses on gathering the system requirements to achieve a fully operational software and the evaluation of requirements gathered. First the stakeholder analysis will be conducted to identify the roles of stakeholders in the system. The methods used for gathering the data required at the evaluation of that data will be carried out next. Along with the evaluation of the gathered requirements a use case diagram is drawn up along with the use case descriptions. Towards the end the functional and non-functional requirements are analysed along with their level of priority.

4.2. Stakeholder Analysis

For the system that we have proposed, we have identified our most important stakeholders to be Lawyers and Law firms. Since the system is proposed to make the lives of the aforementioned easier classifying them as our immediate stakeholders was not a problem. But what about the people who are not entirely interested in the project but will be needing updates on and off? In our analysis we have identified judges and the government as highly interested, but non-immediate stakeholders with regards to the system, (deduced because they will not be the primary target for this systems use but will be concerned with the way how the system works and that it does work with an immense accuracy rate.) but they must be kept informed of the changes made to the system because the law is at work here. Failing to inform them of any new changes could result in a fatal blow for a court case.

As mentioned, Lawyers are our primary stakeholders and thus, we will be closely engaging them with the creation of the system. (interviews with lawyers and explaining to them what the system is, how they would use it, what features they would prefer to have etc.) This will ensure that we satisfy their requirements and that they will provide positive feedback on the system (which will

generate a goodwill towards the system and attract more users). Of course, changes may be required later on with evolving needs, but as before, interviews can be conducted once again to gather more information on what requirements need to be satisfied most.

The stakeholder analysis can be displayed using graphs and the model for our analysis can be found on the following page with appropriate labels. Apart from the aforementioned parties, I proceeded with adding our work force in the project as well since their interest lies in the system. Also you can identify the 'Lawyers Clients' stakeholder I have noted down in the diagram. These are the clients of the lawyers who use the system who can prevent themselves from wasting their resources on an unproductive case. (case where the probability of them winning is quite low)

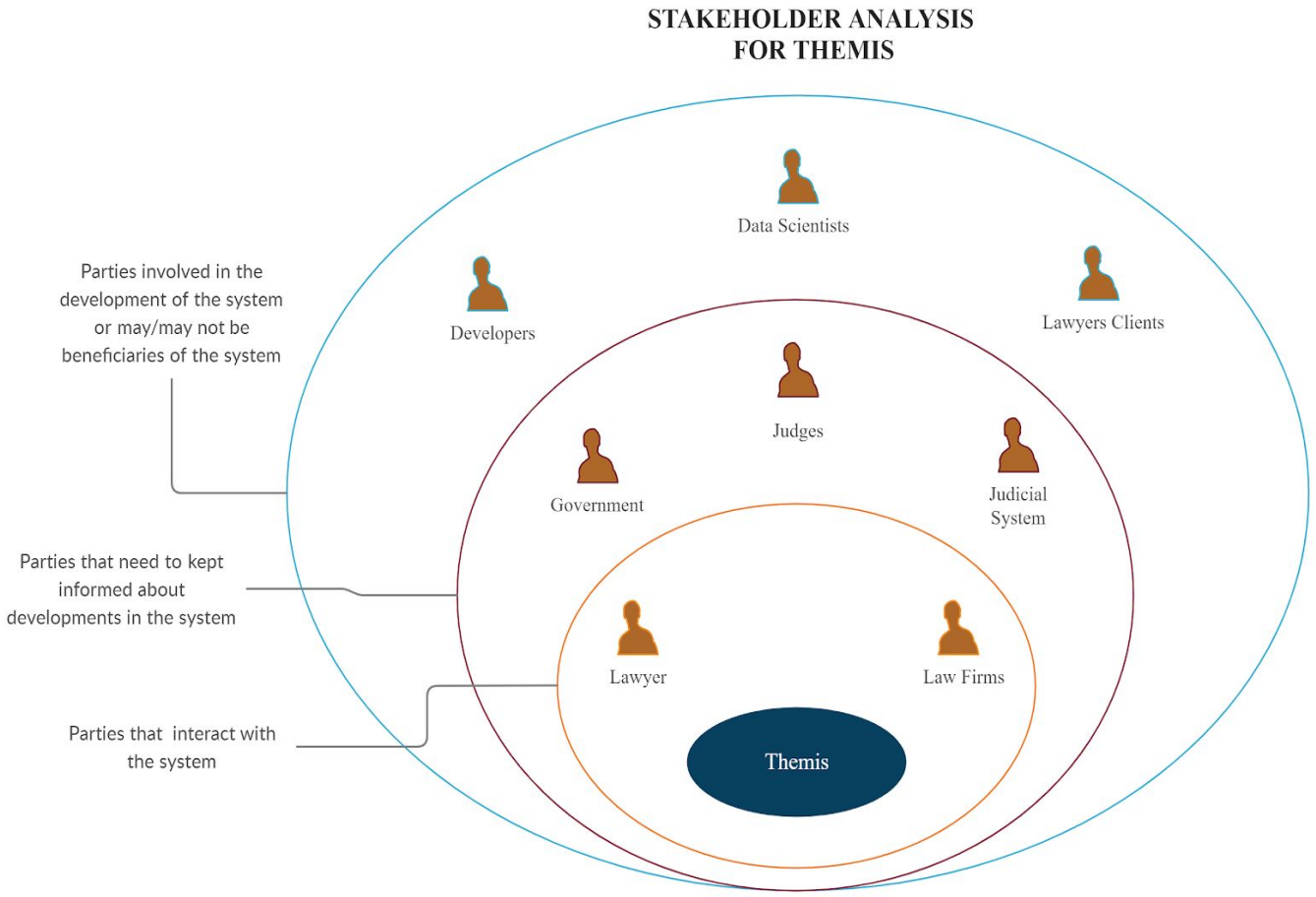


Figure 4.1. Stakeholder Analysis

4.3. Requirement Gathering & Analysis of gathered data

Collection of all requirements is an important step for this research project. Different methodologies were used to collect these requirements and the strengths and drawbacks encountered in each of these methodologies are also discussed here.

4.3.1. Brainstorming Sessions

Brainstorming sessions were conducted many times and in several phases of this project. These were done via group discussions with the team members and involving in a self-analysis process. analysis of the brainstorming sessions used for requirement gathering is stated below.

<table>
<tr><td colspan="2">Findings</td></tr>
<tr><td colspan="2">Several sessions took place with the project supervisors , lecturers , university undergraduates and the team members. During the brainstorming sessions requirements fulfilled by the project , algorithms used , system designs, implementation decisions and the final report were all benefited from the brainstorming sessions.</td></tr>
<tr><td>Advantages</td><td>Disadvantages</td></tr>
<tr><td>Identify requirements that may not have been covered during self research process.

Idea building which includes creative and new ideas to help build the system.

Breaks the routine and helps us reveal improvements we need for the research project as well as a team.</td><td>Input of contradicting ideas and methodologies during the brainstorming sessions made the process a bit difficult.</td></tr>
</table>

Table 4.1 - Brainstorming Sessions

4.3.2. Literature Review

The Literature review of this project has been broken down into 3 main categories to cover every aspect of this research project. The literature review of the domain , a literature review of the technologies used and the literature review of the existing solutions are the 3 main categories in the Literature review. Analysis of the Literature review is as follows.

Findings	
It would be accurate if it is stated that the majority of the contribution towards the data gathering and analysis process was fueled by the Literature review of the research project. Much of the requirements engineering process was immensely benefited from the literature review.	
Advantages	Disadvantages
Benefit of identifying well evaluated and documented findings related to our research project area. Ability to collect a broad range of resources to help and support our research project.	It was time consuming and sometimes lead to a tendency to overlook important “grey literature”.

Table 4.2 - Literature Review

4.3.3. Interviews

Interviews were conducted with domain experts. Interviews helped in broadening our knowledge about the requirements gathered as well as helped to identify more requirements. Several domain experts were interviewed. Mr. Anura Senevirathne , Mr. U.LG Bandara were selected for the interviewing process and they are currently practicing lawyers in Sri Lankan courts.

Findings	
The domain experts were very interested in the system proposed and appreciated the attempts taken to automate some aspects of the legal system. But they also pointed out issues that could arise in this system especially due the breach of independence and transparency of the legal system due to various influences such as political influences. Discussing these issues helped us address the needs of our system to provide satisfying and accurate results.	
Advantages	Disadvantages
Gave us the ability to analyse and classify a judgement more deeply with understanding. Was very efficient	It was difficult to set up the interviews. Not all interviewees had the same thoughts and ideas. Some of it were contradictory.

Table 4.3 - Interviews

4.3.4 Distributing Questionnaires

We distributed questionnaires which was made available through google forms.We used this forms to bring forward specific questions to the target audiences and identify requirements.
This section will be dedicated to discuss the findings, advantages and disadvantages encountered during this requirement gathering process.

Advantages	Disadvantages
It was a good way to meet the target audience. Time saving Analysis of the results can be easily done	Effort put into answering questions in the form is linked with the accuracy of the requirement identification. Difficulty in getting responses from

using the analysis functions made available with the tool.	participants for open ended questions.

Table 4.4 - Questionnaires

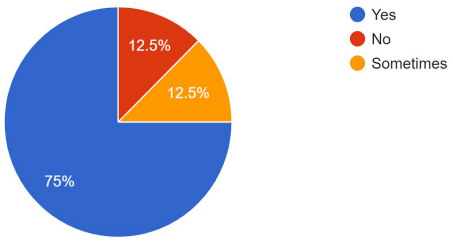
Question	Do you think Verdicts given in a previous judgement is important and must be considered when working on a case?
Aim of question	To identify the importance of verdicts when working on a client case.
Observations	
 Figure 4.2. Observation 1 It was observed that 75% of the respondents agreed that verdicts are important and must be considered when working on a case and however 12.5% of the respondents stated that they don't consider verdicts important in a case.	
Conclusion	
Since the questionnaire was filled by law students it is observed that they have the domain knowledge and all responses are valid and accurate. This question was asked to monitor the importance and the vital role our system will be playing in the domain of study.	

Table 4.5 - Observation 1

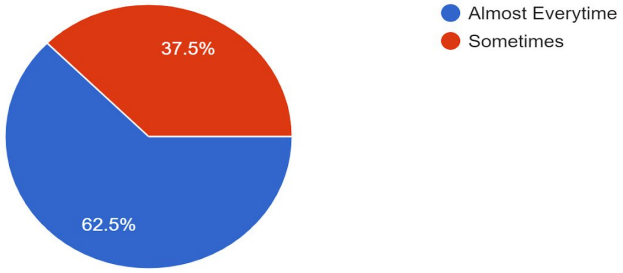
Question	How often do you think you will use Verdicts when dealing with a case?
Aim of question	To identify the use of verdicts when working on a client case.
Observations	
 Figure 4.3. Observation 2 It was observed that more than 60% of the respondents agree that they use verdicts frequently for the cases they work and there is a 37.5% of the respondents that use a verdict at least sometimes when dealing with a case.	
Conclusion	
Verdicts are used rapidly and excessively by lawyers and law students when working on cases and proves that it plays a vital role in a lawyers profession.	

Table 4.6 - Observation 2

Question	How many verdicts do you think you would normally look for to support your research on the case?

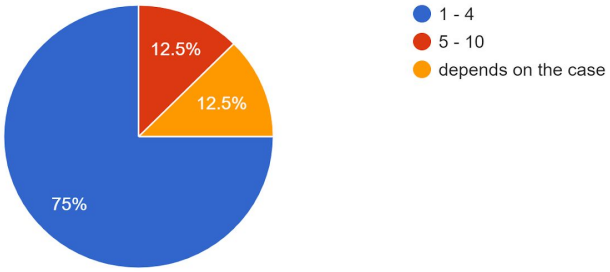
Aim of question	To identify on average how many verdicts are used per case.
Observations	
 Figure 4.4. Observation 3 It was observed that 75% of the participants use approximately 1 to 4 verdicts to support their case while there is a minority that says the amount of verdicts used depends on the case and the other says that they use more than 4 verdicts to support their case.	
Conclusion	
There is very good evidence that per case there are at least more than on verdict used to as support. This once again proves the vital role a verdict plays when dealing with a legal case.	

Table 4.7 - Observation 3

Question	How long do you normally spend researching for verdicts that help your case?
Aim of question	To identify how much time is currently spent in finding supporting verdicts.
Observations	

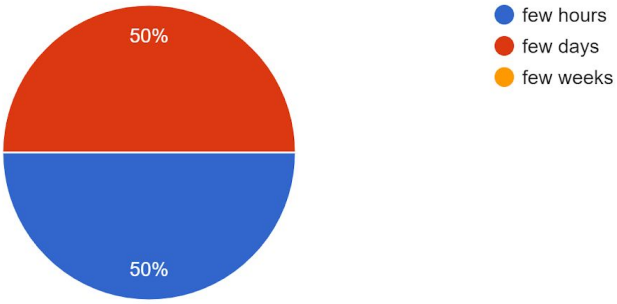


Figure 4.5. Observation 4

It was surprising to see that there was exactly one half of participants who spend a few days researching for the verdicts while the other half would spend a few hours to get the verdicts. However there was no participant who spent a few weeks working on finding proper verdicts.

Conclusion
All the participants at least spend a few hours or days when searching for supporting verdicts. This shows that our system can immensely benefit them by suggesting them the supporting verdicts for the case they are working on in just a few minutes.

Table 4.8 - Observation 4

Question	Do you think there is a chance of missing a good supporting verdict to your case in the process of your research?
Aim of question	To identify if there is a possibility of missing any important verdicts using the current methods they follow.
Observations	

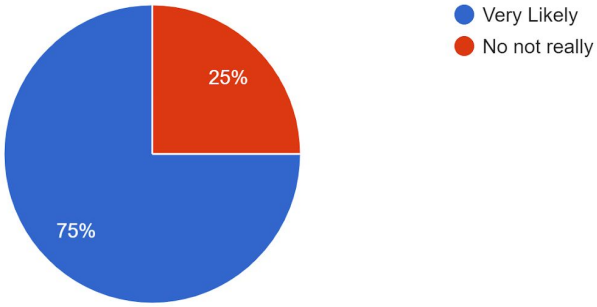


Figure 4.6. Observation 5

Observations show that 75% of the participants agree that using traditional methods of looking for verdicts has a high chance of missing a verdict that is supporting a case they are working on.
Conclusion
Majority of the participants declared that there is a possibility of missing an important verdict on the process of the research for their case. This shows that our system can immensely help them to assure that no similar verdict is missed for the case they are working on.

Table 4.9 - Observation 5

Question	Would it be helpful if there was a system that predicts the probability of winning the case or research you are working on based on previous verdicts available?
Aim of question	To identify if the participants agree that it would be helpful for them if our system can give a prediction of the probability of winning the case they are working on based on previous verdicts available.

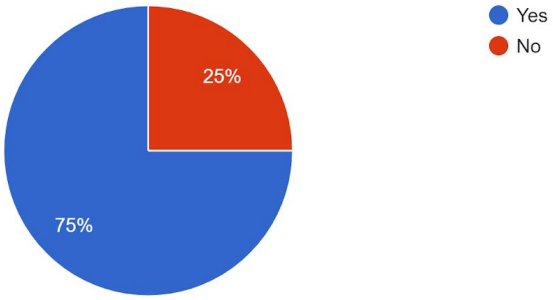
Observations
 Figure 4.7. Observation 6 It was observed that majority of the participants agreed that it would be helpful if the system can predict the probability of winning the case they are working on based on previous verdicts given. A 25% of the participants however disagree with this.
Conclusion
Predicting the probability of winning a case based on previous verdicts will help lawyers get more efficient and straightforward decisions when accepting client cases. Before dealing with the case they can seek the assistance of our system to decide the probability of winning this case as then decide if they are willing to take the case or not.

Table 4.10 - Observation 6

Question	How convenient would it be if you had a system that helps to list you all the identical case judgments and related verdicts you can use for a particular case or research you are working on?

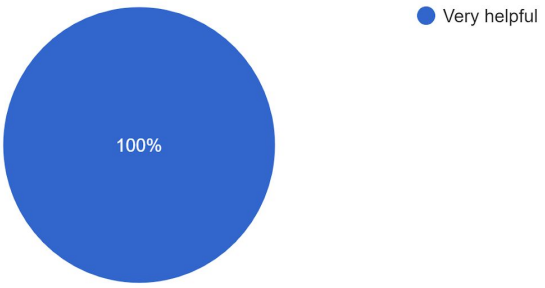
Aim of question	To identify if the participants are willing to embrace and try our system.
Observations	
 Figure 4.8. Observation 7 It was observed that all the participants are willing to try out our system to help them in their research processes even though we expected more 'other' responses.	
Conclusion	
All the participants believed that it would be convenient to have a system that helps them list all identical case judgments and related verdicts for a particular case they are working on. Also proving the fact that they would prefer to try our system instead of the traditional method of looking for cases and verdicts.	

Table 4.11 - Observation 7

Question	Do you think it would be helpful if the system suggested you the most likely verdict that can be delivered for the case you are working on?

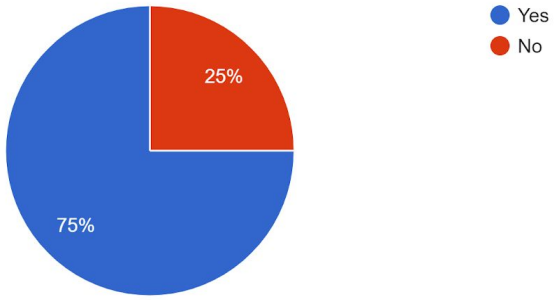
Aim of question	To identify if the participants agree that using our system that predicts the most likely verdict to be delivered on the case they are working on will help them.
Observations	
 Figure 4.9. Observation 8 It was observed that majority of the participants prefer the system giving them a prediction of the most likely verdict to be given on the case they are working on. While there is a 25% of the participant group who stated that they dislike the feature.	
Conclusion	
Predicting the most likely verdict to be delivered on a case they are working is indeed a very helpful feature in our system. But however there are valid reasons as to why some may disagree to this. A prediction given by our system is not 100% accurate. If they totally rely on the system there can be a slight chance that they may have a wrong prediction on their hands. However as professionals using our system to get an idea and using their skills and experience can make the process efficient as well as accurate.	

Table 4.12 - Observation 8

4.3.5 Summary of Findings

Finding	Literature Review	Questionnaire	Brainstorming session	Interviews
Importance of verdicts in a law case.	✔	✔	✔	✔
Importance of using accurate and reliable datasets	✔		✔	✔
Usage of classification algorithms to build the system	✔		✔	
Usage of regression algorithms to build the system	✔		✔	
Importance of attractive and user friendly graphical user interface.			✔	✔
Importance of reliability of system by user.	✔	✔	✔	✔
Importance representing computed information in a clear and understandable way to the user			✔	✔
Ability of system to only allow users of legal profession to log in to the system			✔	✔

Importance of focusing on the accuracy of the results generated	✔		✔	✔

Table 4.13 - Summary of Findings

4.3.6 Context Diagram

A concept diagram is depicted to clarify and define the boundaries of the software system and note down the internal and external components before development.

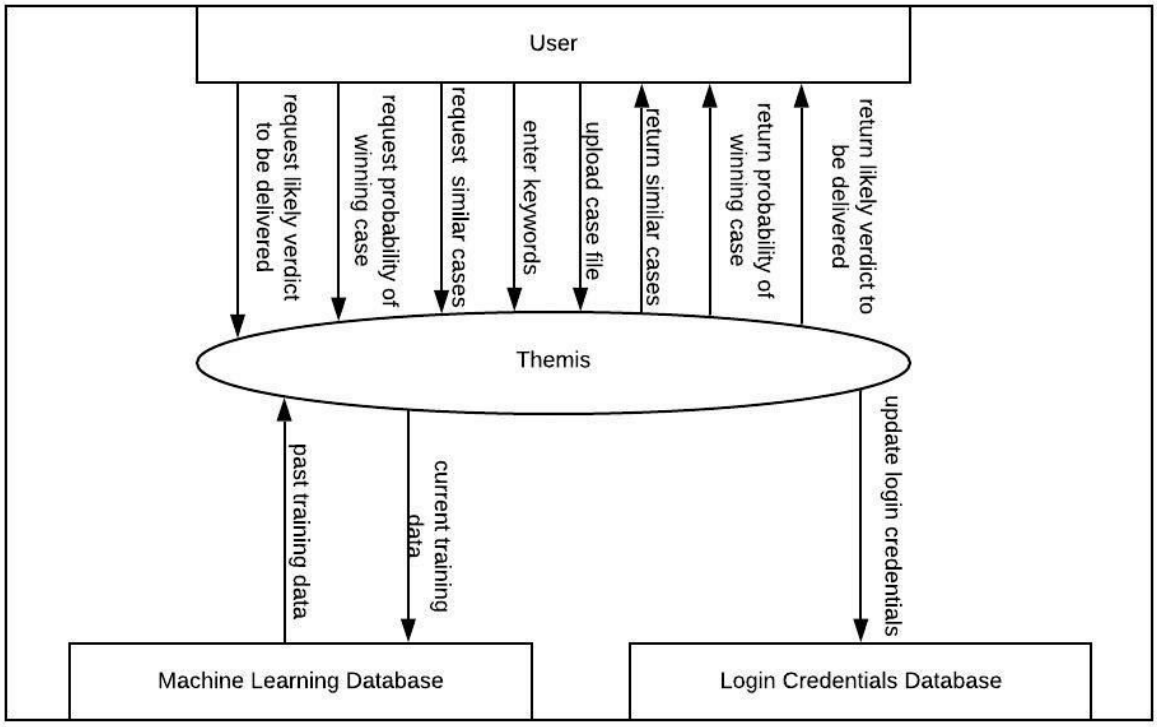


Figure 4.10. Context Diagram

4.4 Models

4.4.1 Use Case Diagram

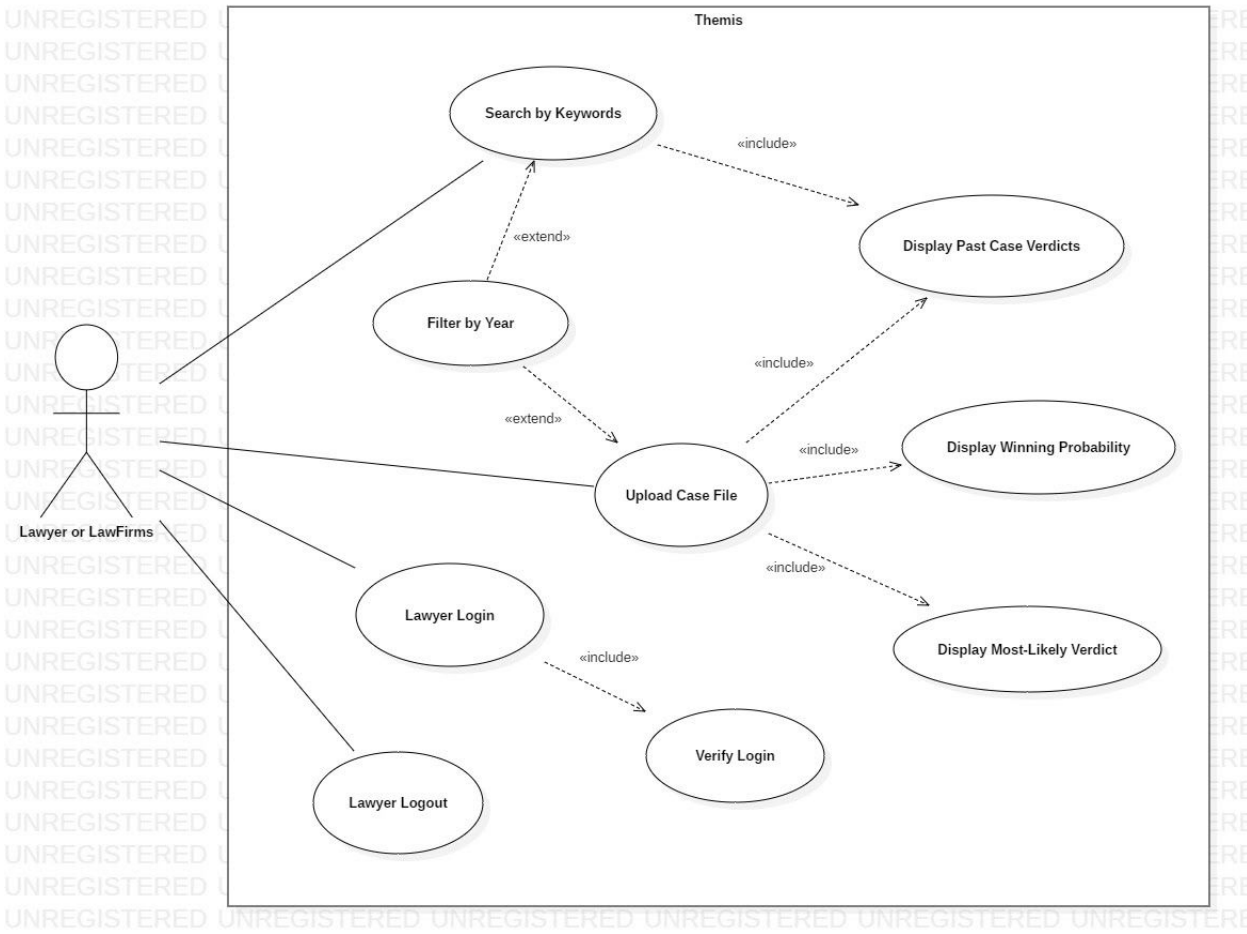


Figure 4.11. Use Case Diagram

4.4.2 Use Case Description

Use Case Name	**Upload case file**
Description	Searching the similar case verdict files based on the case file submitted.
Participating Actors	Lawyers/Law Firms
Pre-Conditions	The user should have the case details.
Extended Use Cases	Filter by year
Included Use Cases	Display past case files. Display winning probability. Display verdict.
Main Flow	● User chooses to search for similar case verdict files ● User upload current case file
Alternative Flow	None
Exceptional Flows	If the application shuts down during this process the use case ends in failure
Post-Conditions	System will display all the similar case verdicts, probability of success and verdicts found related to the case files entered by the user.

Table 4.14 - Use Case Description 1

Use Case Name	**Display past case verdicts**
Description	Display the similar case verdicts based on the case details or case file submitted.
Participating Actors	Lawyers/Law Firms
Pre-Conditions	The user should have the case details or case file.
Extended Use Cases	None

Included Use Cases	None
Main Flow	● User Enters Case Details or current case file to search for similar case verdicts
Alternative Flow	None
Exceptional Flows	If the application shuts down during this process the use case ends in failure
Post-Conditions	System will display all the similar case files found related to the case details entered by the user.

Table 4.15 - Use Case Description 2

Use Case Name	**Display winning probability.**
Description	Display the probability of the given case file based on the success rate of similar cases.
Participating Actors	Lawyers/Law Firms
Pre-Conditions	The user should have the current case file.
Extended Use Cases	None
Included Use Cases	None
Main Flow	● User Enters Case file to find the probability of the case success.
Alternative Flow	None
Exceptional Flows	If the application shuts down during this process the use case ends in failure
Post-Conditions	The system will calculate and display the probability of the case success with the provided case details.

Table 4.16 - Use Case Description 3

Use Case Name	**Display most-likely verdict.**
Description	Providing the user with the most-likely case verdict that will be delivered based on previous similar cases and the country's laws.
Participating Actors	Lawyers/Law Firms
Pre-Conditions	The user should have the case details.
Extended Use Cases	None
Included Use Cases	None
Main Flow	● User Enters Case file to see the most likely case verdict.
Alternative Flow	None
Exceptional Flows	If the application shuts down during this process the use case ends in failure
Post-Conditions	The system will provide the user with the most probable case verdict that can be delivered for the given case details.

Table 4.17 - Use Case Description 4

Use Case Name	**Search by keywords**
Description	Searching the similar case verdict files based on the case keywords and filters user submitted.
Participating Actors	Lawyers/Law Firms
Pre-Conditions	The user should have the current case file.
Extended Use Cases	Filter by year
Included Use Cases	Display past case files.
Main Flow	● User chooses to search for similar case verdict files ● User enter case details using specific keywords

Alternative Flow	None
Exceptional Flows	If the application shuts down during this process the use case ends in failure
Post-Conditions	System will display all the similar case verdicts related to the keywords user entered.

Table 4.18 - Use Case Description 5

Use Case Name	**Lawyer login**
Description	Lawyers or law firms enter their identification number, email address and password in order to get access to the system.
Participating Actors	Lawyers/Law Firms
Pre-Conditions	The user should have the legal identification number and correct details.
Extended Use Cases	None
Included Use Cases	Verify login
Main Flow	● User chooses to login. ● User enter identification number ● User enter email address ● User enter password
Alternative Flow	None
Exceptional Flows	If identification number is invalid user should re-enter. If the application shuts down during this process the use case ends in failure
Post-Conditions	System will display the home page.

Table 4.19 - Use Case Description 6

Use Case Name	**Verify login**
Description	System search whether the input details are valid or not.
Participating Actors	None
Pre-Conditions	The user should have the legal identification number and correct details.
Extended Use Cases	None
Included Use Cases	None
Main Flow	● System verify user identification number ● System verify email address
Alternative Flow	None
Exceptional Flows	If identification number is invalid user reload the login page. If the application shuts down during this process the use case ends in failure
Post-Conditions	System will display the home page.

Table 4.20 - Use Case Description 7

Use Case Name	**Filter by year**
Description	Searching the similar case verdict files based on the case held year.
Participating Actors	None
Pre-Conditions	The user should give the year.
Extended Use Cases	None
Included Use Cases	None
Main Flow	● User Enters year to find similar case verdicts from that year.
Alternative Flow	None

Exceptional Flows	If the application shuts down during this process the use case ends in failure
Post-Conditions	The system will provide the user with the most probable case verdict that can be delivered for the given year and case details.

Table 4.21 - Use Case Description 8

Use Case Name	**Lawyer logout**
Description	Lawyers or law firms click the logout button in order to logout from the system.
Participating Actors	Lawyers/Law Firms
Pre-Conditions	The user should have logged in to the system before logging out.
Extended Use Cases	None
Included Use Cases	None
Main Flow	● User chooses to logout. ● System logout.
Alternative Flow	None
Exceptional Flows	If the application shuts down during this process the use case ends in failure
Post-Conditions	System will display the login page.

Table 4.22 - Use Case Description 9

4.4.3 Domain Model

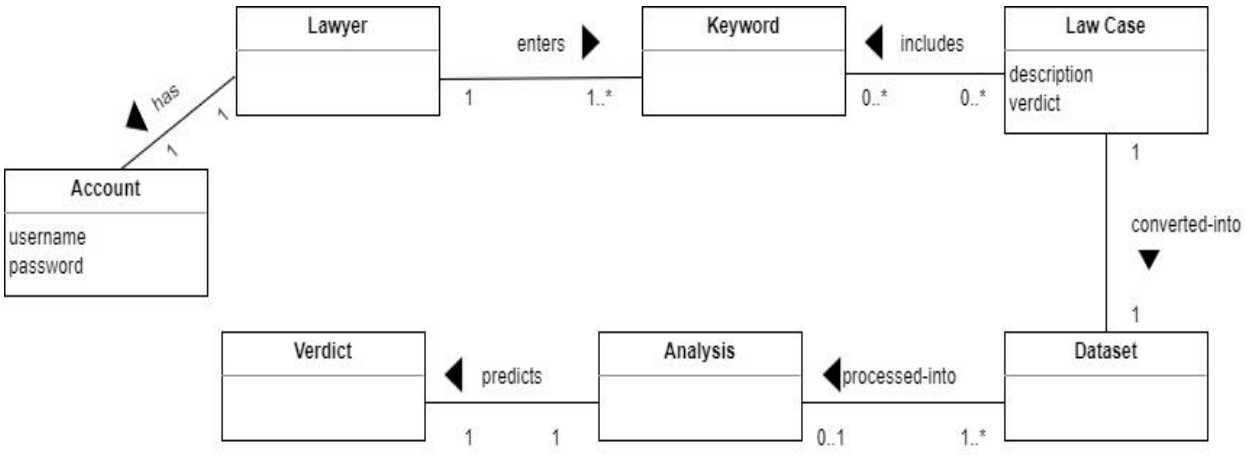


Figure 4.12. Domain Model

This Domain Model helps us to identify the key components of the application and it also describes the relationships between each of those components. By looking at this diagram we can understand that when a lawyer enters a keyword(s) that include in law cases which have been converted into datasets are processed into an analysis in order to get a most likely verdict. So this diagram is important in having a basic idea of how the system's components behave and also helpful to find out the object oriented nature of the application.

4.5 Requirement Specifications

This section describes the requirement specifications for the software. It is prioritized according to three levels

Priority Level	**Description**
High	Essential functionality for the software
Medium	Not an essential requirement, but necessary

Low	Out-of-Scope requirements.

Table 4.23 - Priority Level

4.5.1 Functional Requirements

Requirement ID	Requirement	Description	Priority
FRQ01	User login	The system should display a login for lawyers and clients respectively.	High
FRQ02	Display lawyers' features.	The system should provide all the features at the lawyer's disposal once logged in using the lawyers login.	High
FRQ03	Predict verdict	The system should be able to predict the most likely verdict of a given case/scenario.	High
FRQ04	Detect keywords	The system should be able to read a case file and extract necessary keywords for searching.	High
FRQ05	Winning probability	The system should be able to predict the probability of winning or losing a case once given the case details in the form of a percentage.	High
FRQ06	Search by keywords	Lawyers and clients should be able to search previous cases which are related using keywords, filters and the case file.	High

FRQ07	Analysis of previous cases	Lawyers should be able to view an analysis of previous cases.	High
FRQ08	Detect misspelled words	The system should be able to identify incorrect words which user enters	Low

Table 4.24 - Functional Requirements

4.5.2 Non-Functional Requirements

Requirement ID	Requirement	Description	Priority
NFRQ01	Performance	The user should be able to view analysis, previous cases etc. without taking too long.	Medium
NFRQ02	Security	The system should be able to authenticate lawyers and clients. As in, lawyers should log in to view analysis and prediction and clients must log in to view previous cases.	High
NFRQ03	Reliability	All previous cases that are given to the user and used for the prediction and analysis should be certified by the Supreme Court of the Democratic Socialist Republic of Sri Lanka.	High
NFRQ04	Accuracy	The output of the prediction is expected to be over 80% at all times.	High

NFRQ05	User friendliness	The output of the prediction and analysis should be shown in graphs, charts,etc which are user friendly.	Medium

Table 4.25 - Non-functional requirements

4.6 Chapter Summary

This chapter is kicked off with determining the stakeholders and their role in and during this project development. The different methods for gathering requirements were analysed and a use case diagram to sum up this evaluation was drawn up. Towards the end, Functional and Non-Functional requirements were stipulated.

5.Design

5.1. Chapter Overview

This chapter focuses on the design segment of this project. It features the following diagrams: High Level Architecture Diagram, Class Diagram, Activity Diagram, Sequence Diagram,UI Wireframe Diagram. Along with diagrams are the reasoning as to why these designs were chosen to be the final.

5.2 Design Goals

Below are the design goals which have been set for the implementation of the project.

Design Goal	**Description**
Accuracy	This system aims to provide the user with the most accurate results based on the input given to the system by the user. The system would not be considered as accurate if the users are not able to utilize the features and attain plausible results.
Adaptability	The system should be able to adapt to new elements which would be added over time. Changes to the existing system should be easy to make and should not cause the entire system to collapse during the changing process.
Performance	Results should be provided under a short period of time. Computational potential should be used efficiently to provide deft and diligent performance.
Scalability	Taking the large datasets that will be used for the system, the system should be able to handle the weight and be capable of scaling into further heights without much hassle.

Reusability	The system units should be reusable in other developments and thus should be built with flexibility in mind.

Table 5.1 - Design Goals

5.3. High Level Architecture Diagram

This diagram consists of three different layers. The presentation layer which focuses on the interfaces that are provided to the user (Lawyers/ Law firms) for their usage. The logic layer consists of the business logic and the functions that are related to those which are available to the user. At the foundation of the architecture diagram is the Data and Engine Layer which consists of the engines that work with the database.

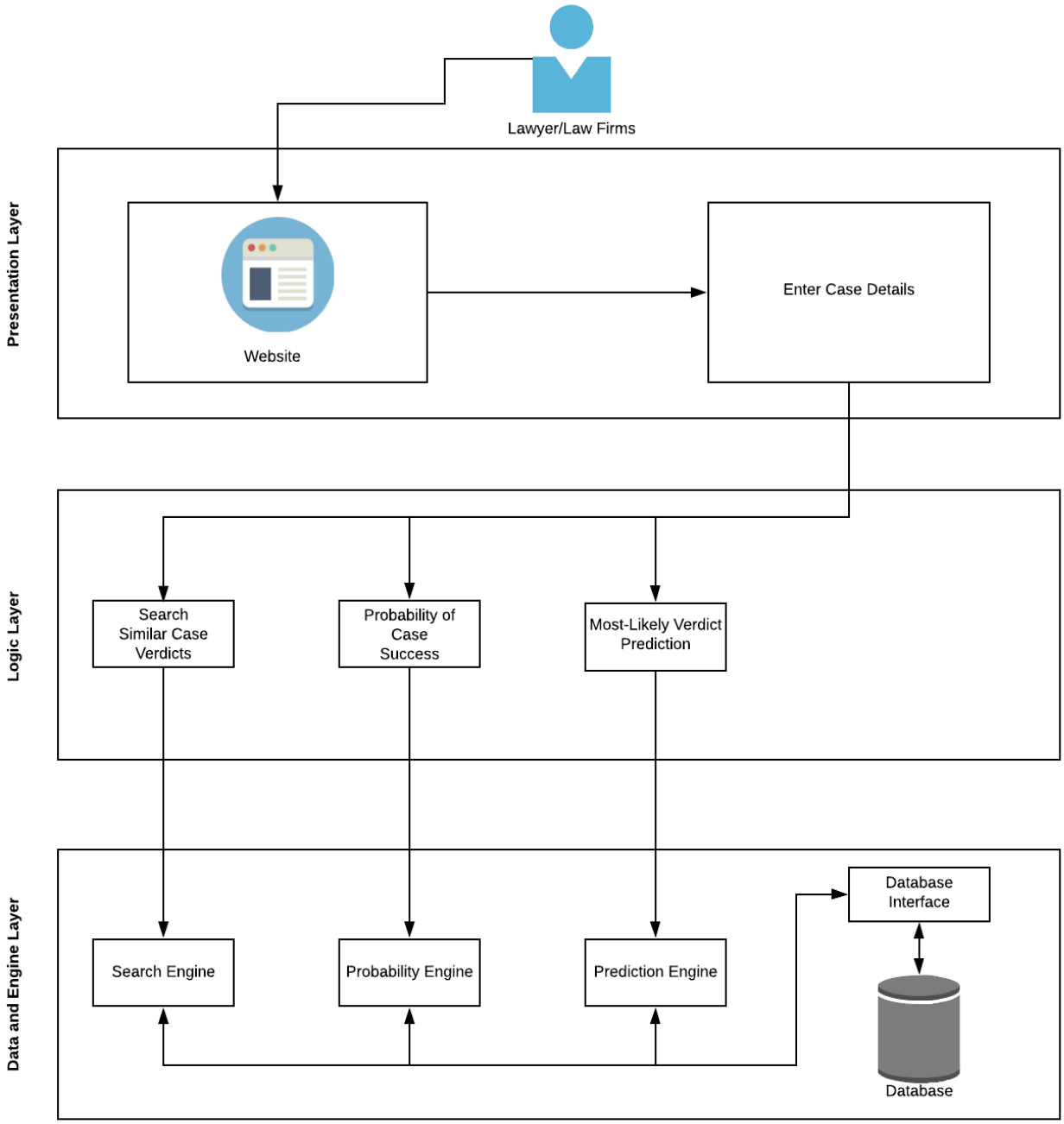


Figure 5.1 High Level Architecture Diagram

5.4. Class Diagram

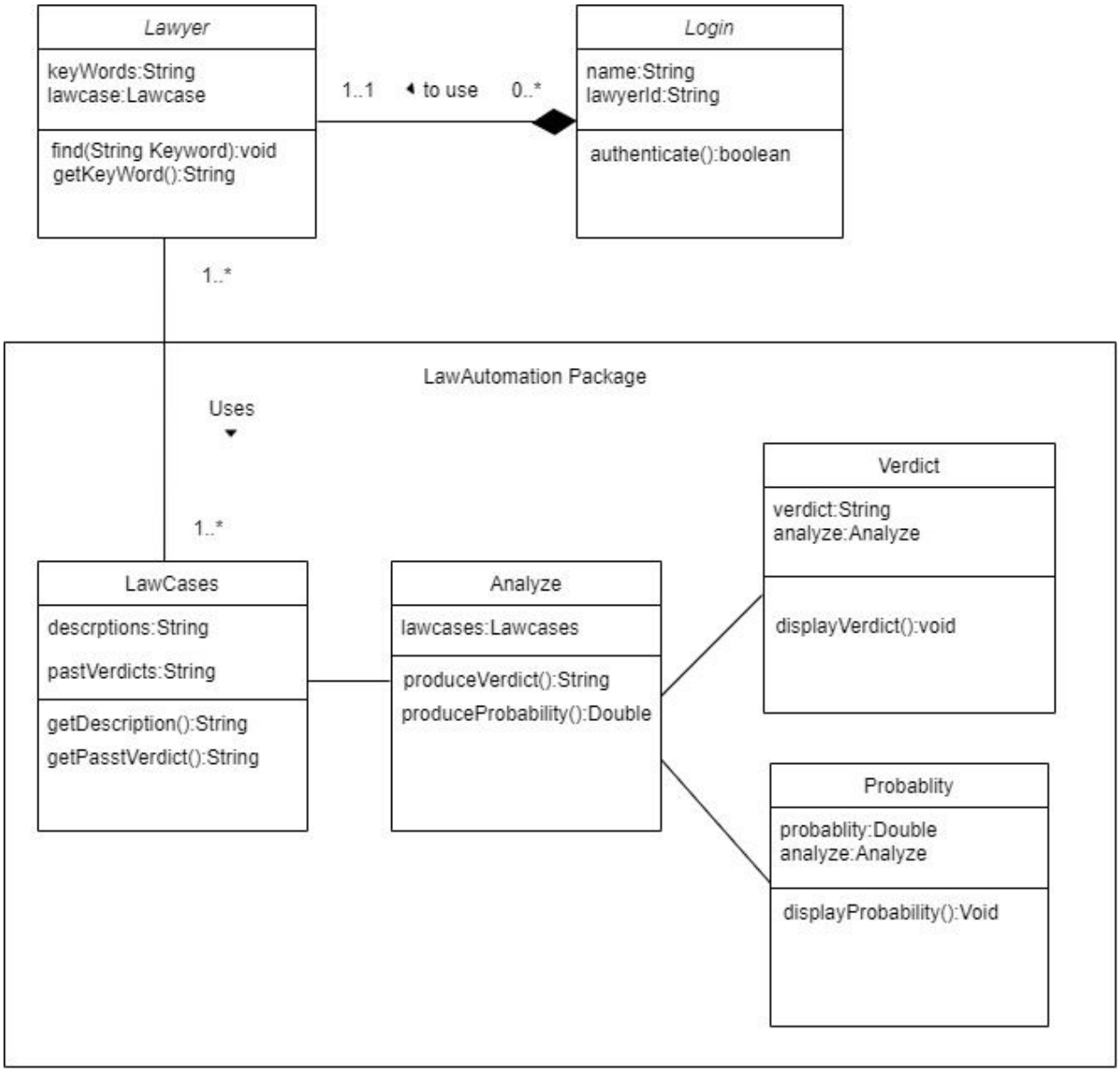


Figure 5.2. Class Diagram

This section follows the classes that will be developed in the system. As in the diagram for a lawyer to use this system, they have to login in to the system, where he/she has to enter their respective id to verify whether they are lawyers then the class lawyer collects/captures the

keywords entered by a lawyer(user). And then the lawCases class will use the keywords in Lawyer to find similar cases and display them while collecting their verdicts and results. Then the analysis class analyze the data. Then the verdict class and the probability class will use the analysed data to produce the possible verdict and the percentage/probability of the win of the case. So this class diagram explains the object oriented nature of the system.

5.5. Activity Diagram

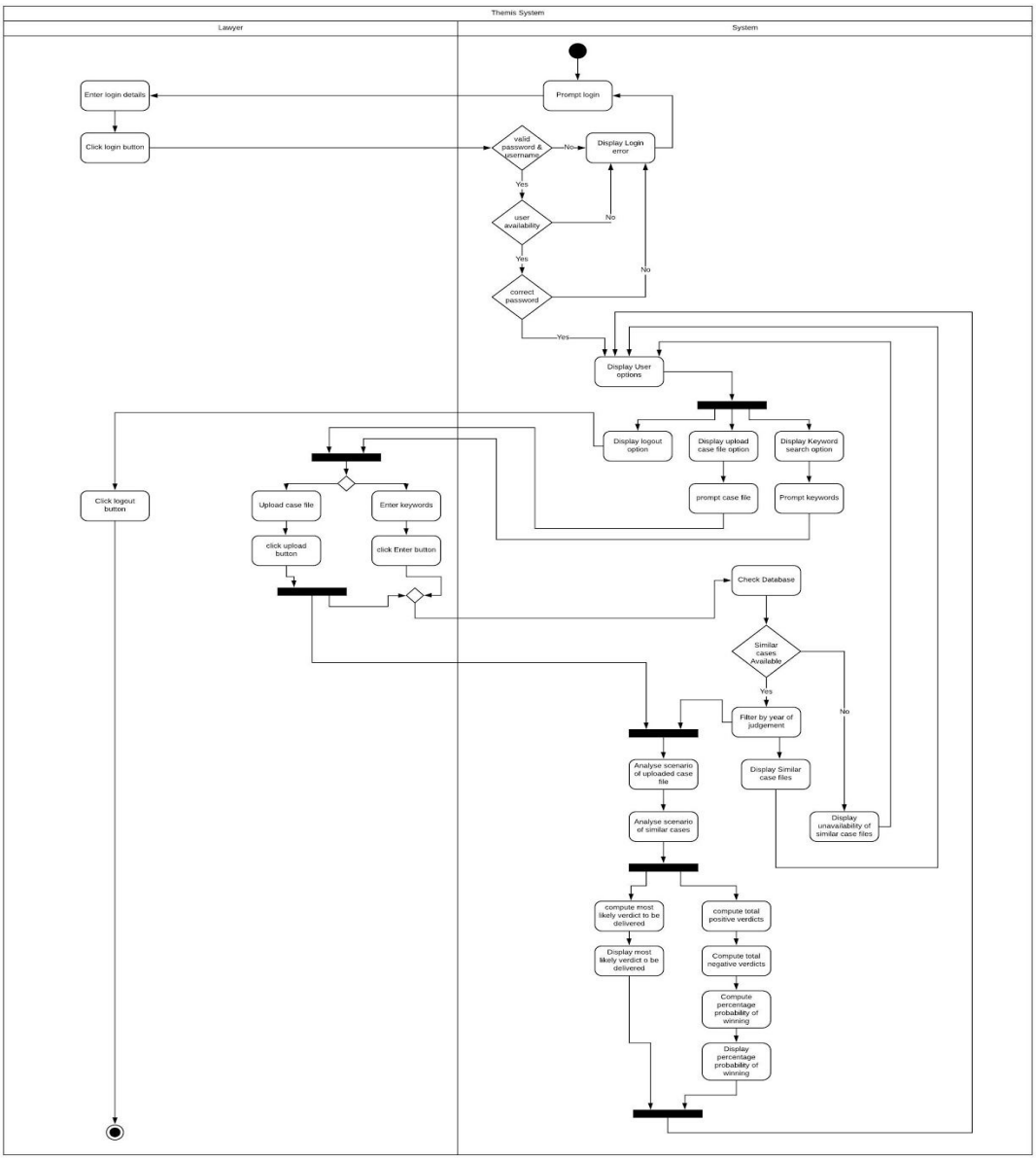


Figure 5.3. Activity Diagram

The activity diagram depicts and illustrates the control flow and the behavior of the Themis system in reference to the use cases provided in the use case diagram. It shows the sequential flow of the system starting from the user login and also shows the concurrent activities that occur in the system such as when the user chooses the upload case file option.

Please refer to the Appendix B of the report for a clear representation of the activity diagram.

5.6. Sequence Diagram

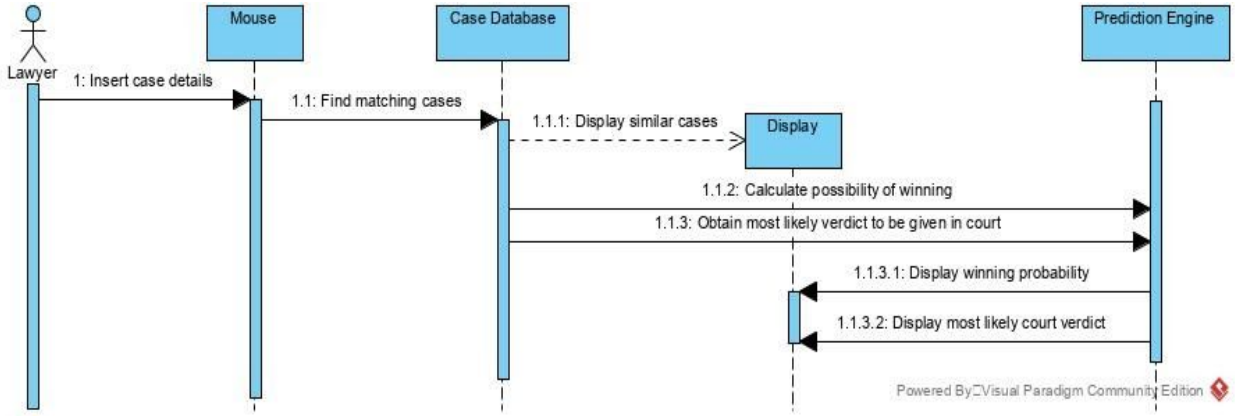


Figure 5.4. Sequence Diagram- Entire System

When compared with the Use case and High level architecture diagrams we can see that the users interaction with one action triggers more events to take place. And in accordance with that our Sequence diagram suggests the same. The lawyer/law firms upload their case details into the site and the case is analysed for keywords and based upon that, the search results of similar cases are retrieved and displayed. Along with that, the probability of winning the case is generated along with the most likely verdict as a reply to the case data being uploaded. This sequence diagram displays the sequence of the system as a whole, further breakdowns are shown below. In the case that any of the sequence diagrams may not be visible, they have been placed in the Appendix A as well.

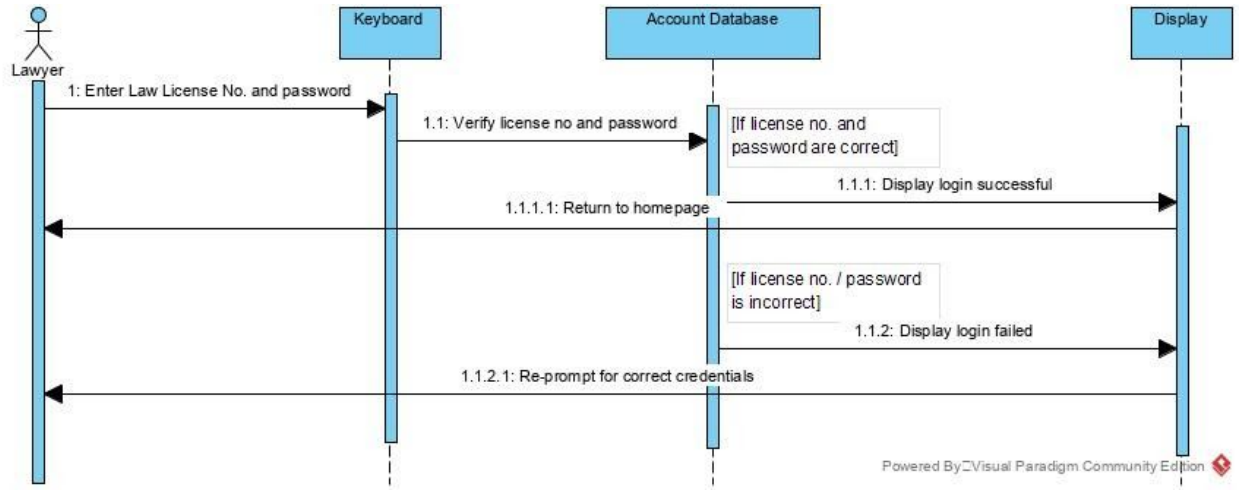


Figure 5.5. Sequence Diagram - Login Lawyer

In the above diagram I depict the sequence of actions that take place when the user decides to login to the website, after providing his/her law license no and password the database is searched for the number and password and logs the lawyer into the system should his/her license no. and password prove to be correct. And will be re-prompted if it is invalid.

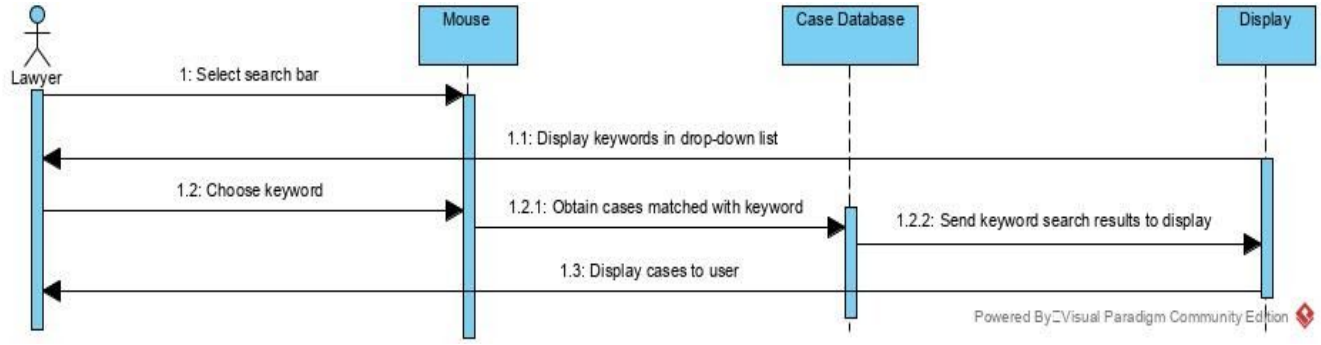


Figure 5.6. Sequence Diagram - Search by Keyword.

As shown in the use-case diagram, the search by keyword functions sequence of actions are displayed in the above diagram. Once the user clicks the search bar, a drop-down list displays all the keywords by which we have sorted the cases, and will return all cases that match the keyword the user chooses.

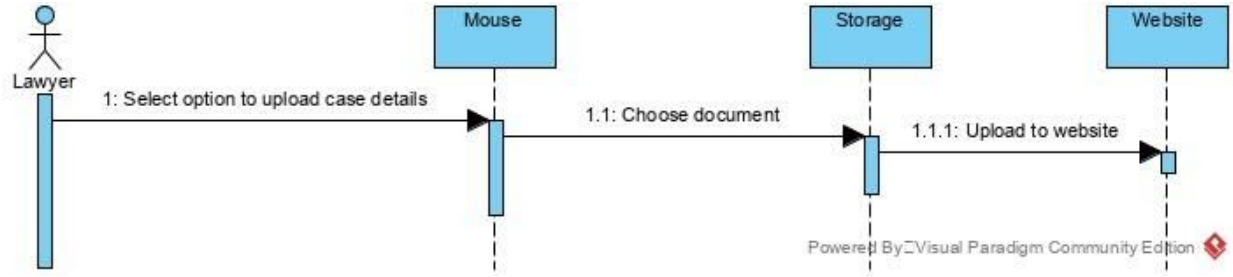


Figure 5.7. Sequence Diagram - Uploading a Case File

The user will have the option to upload the case details to the site by choosing the file from his local storage. The file he/she chooses will then be uploaded to the website.

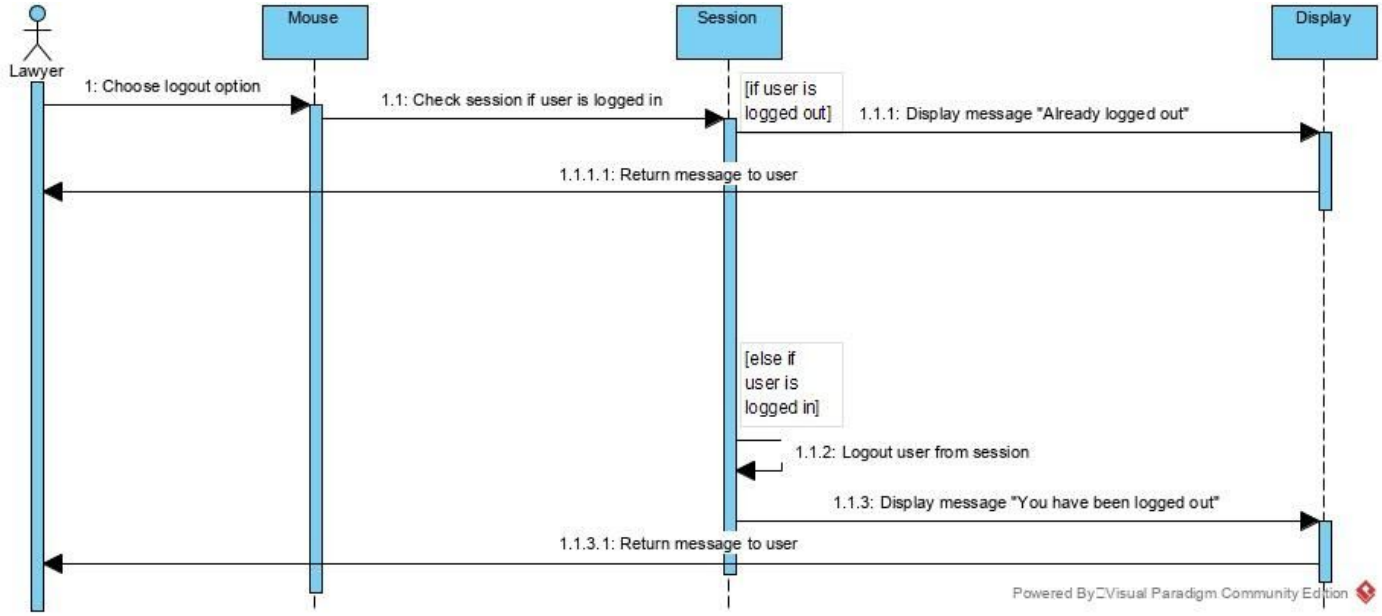


Figure 5.8. Sequence Diagram - Logout from System

If the user has the ability to login, he/she must have the option to logout as well and thus, the logout function checks the login status of the user and performs the relevant actions based on that. (as shown in the diagram, the user is shown different messages based on whether he is online/offline)

5.7. UI Wireframe Diagram

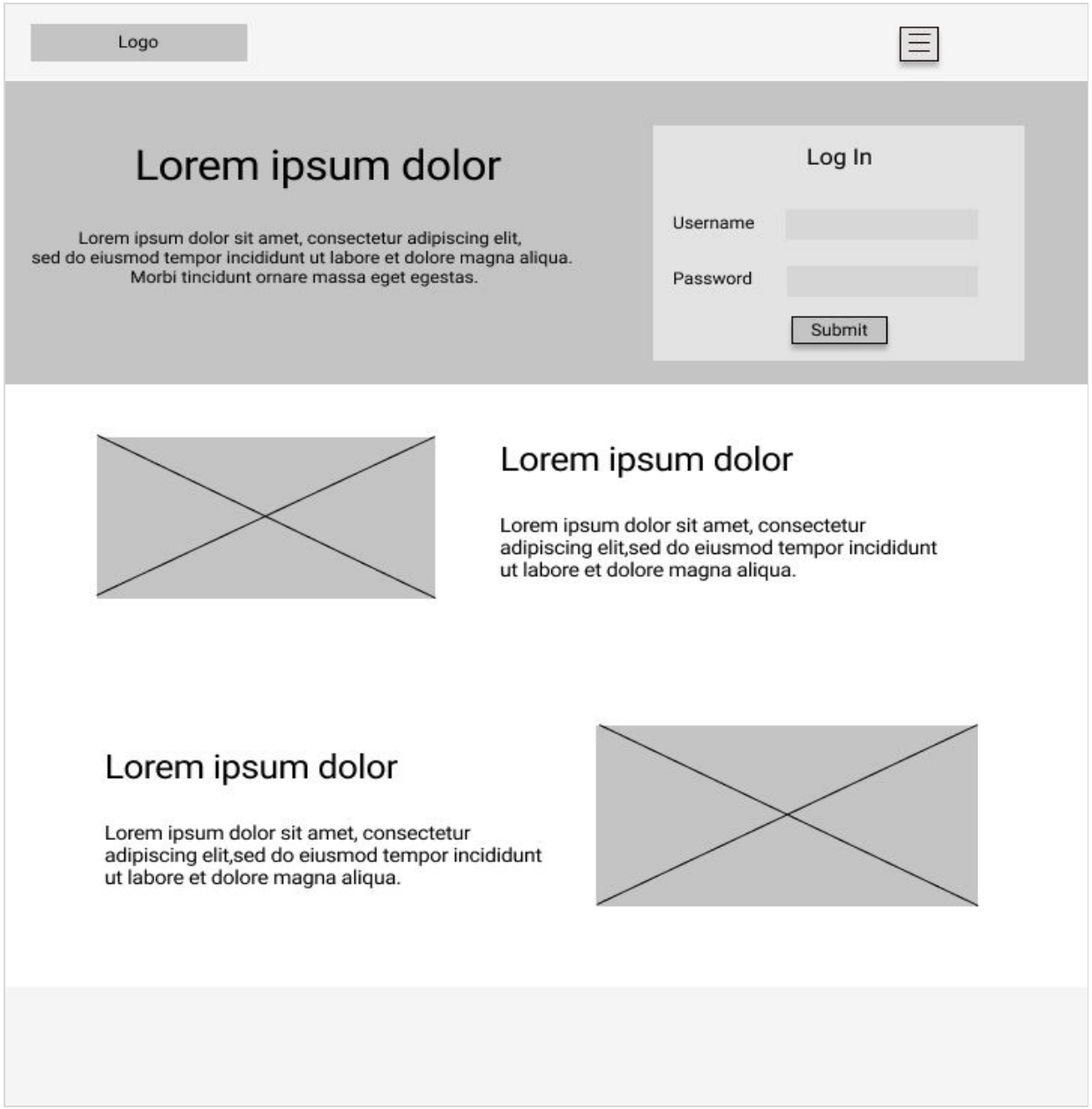


Figure 5.9. UI/UX Wireframe Diagram 1(Home page)

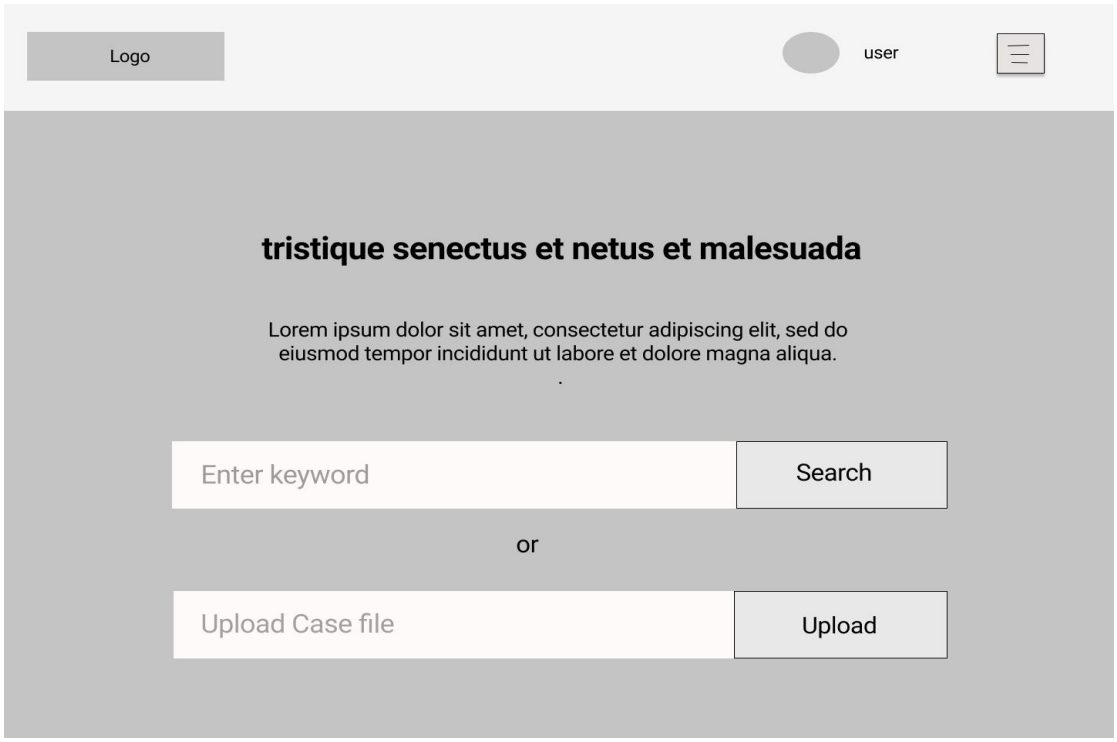


Figure 5.10. UI/UX Wireframe Diagram 2

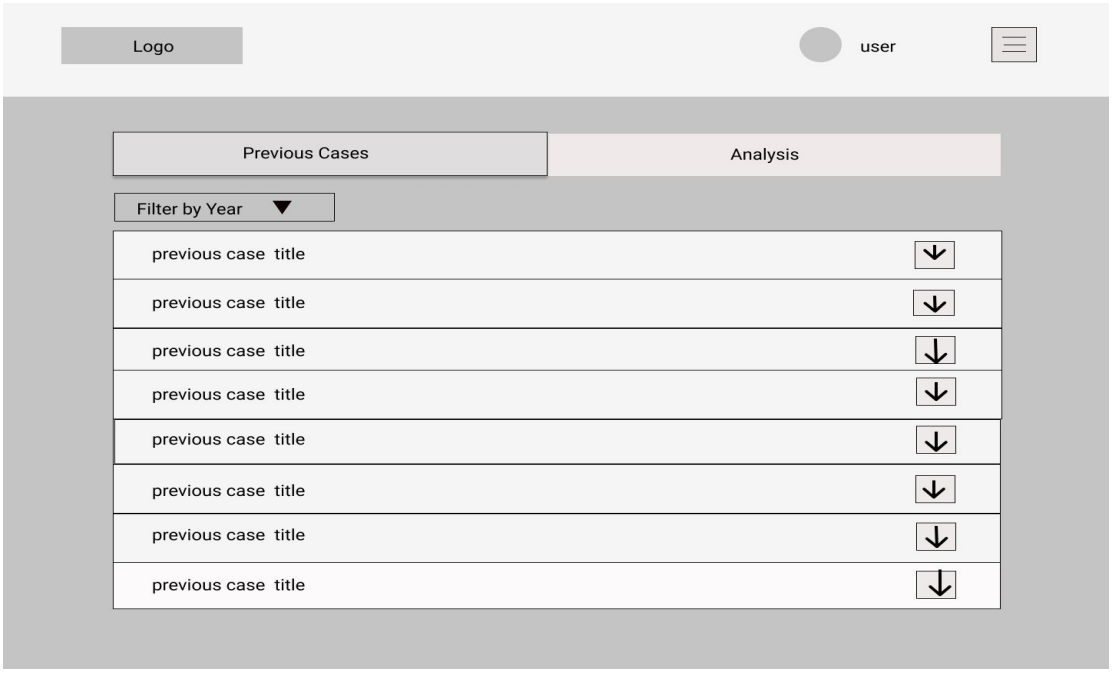


Figure 5.11. UI/UX Wireframe Diagram 3

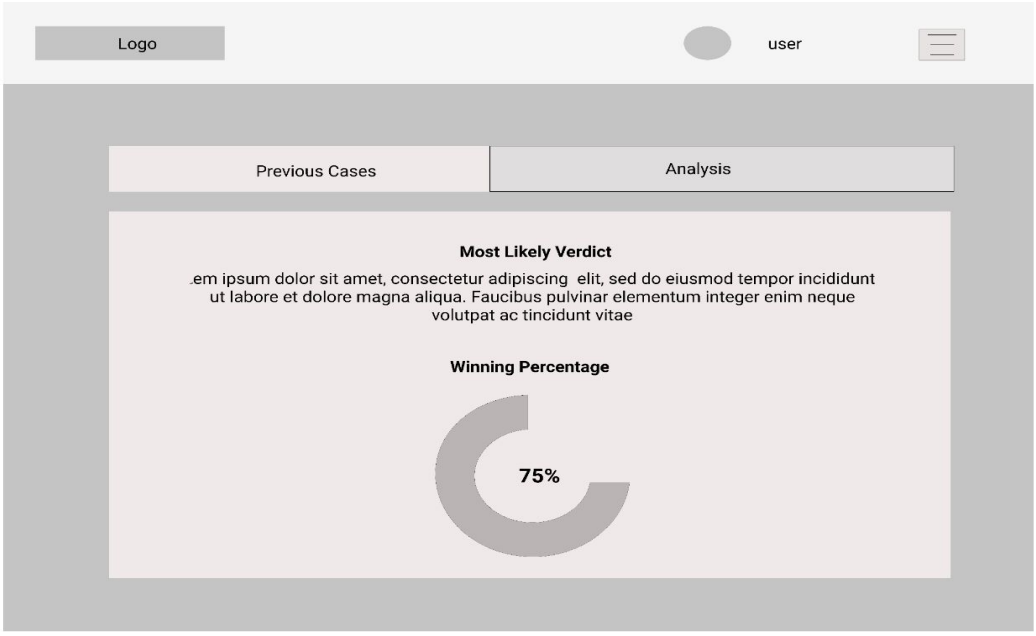


Figure 5.12. UI/UX Wireframe Diagram 4

The above wireframes have been created as a visual guide that signifies the structure of web pages and explains the user's interaction in this application. Our 'Themis' application is a web site that consisted of two main pages which are the home page and the user page. The home page consists of a login form and few other details about the system. Once someone successfully logs in, he or she will be directed to their user page. Then the user will have the chance to search for similar cases, get a most likely verdict and a winning probability by entering keywords or uploading a case file as shown in the second wireframe. Then the corresponding results will appear as shown in the third and fourth wireframes.

5.8. Chapter Summary

This chapter commenced with the High Level Architecture Diagram and its rationale. It went on to show the Class, Activity, Sequence Diagrams and their respective reasoning. Finally the UI Wireframe diagram and its justification brought this chapter to a close.

6.References

lawgovpol.com. (2019). *Case study: Donoghue v. Stevenson (1932)*. [online] Available at: https://lawgovpol.com/case-study-donoghue-v-stevenson-1932/ [Accessed 5 Nov. 2019].

Medvedeva, M., Vols, M. & Wieling, M. Artif Intell Law (2019). https://doi.org/10.1007/s10506-019-09255-y

Katz, Daniel Martin et al. "A general approach for predicting the behavior of the Supreme Court of the United States." *PloS one* vol. 12,4 e0174698. 12 Apr. 2017, doi:10.1371/journal.pone.0174698

Medium. (2019). How Law is being transformed by Machine Learning. [online] Available at: https://medium.com/@giacaglia/how-law-is-being-transformed-by-machine-learning-3642045a0d34 [Accessed 19 Nov. 2019].

Anon, (2019). [online] Available at: https://www.slideshare.net/BhaskerGupta1/predicting-outcome-of-legal-case-using-machine-learning-algorithms-by-ankita-singh-service-delivery-s [Accessed 19 Nov. 2019].

Rayo, E. (2019). *AI in Law and Legal Practice – A Comprehensive View of 35 Current Applications | Emerj*. [online] Emerj. Available at:

https://emerj.com/ai-sector-overviews/ai-in-law-legal-practice-current-applications/ [Accessed 19 Nov. 2019].

Anon,(2019).[online]Available at: https://www.researchgate.net/publication/326202085_Scope_of_Artificial_Intelligence_in_Law [Accessed 19 Nov. 2019]

MathWorks, (2017). What Is Machine Learning? [online]Available at: https://in.mathworks.com/discovery/machine-learning.html [Accessed 5 jan.2020].

Diksha Khurana, Aditya Koli1, Kiran Khatter and Sukhdev Singh, 'Natural Language Processing: State of The Art, Current Trends and Challenges', n.d.
IONOS Digitalguide. (2019). *How does natural language processing work?*. [online] Available at: https://www.ionos.com/digitalguide/online-marketing/online-sales/how-does-natural-language-processing-work/ [Accessed 3 Jan. 2020].

(Medvedeva, Masha & Vols, Michel & Wieling, Martijn. (2019). Using machine learning to predict decisions of the European Court of Human Rights. Artificial Intelligence and Law. 2019. 10.1007/s10506-019-09255-y.).

Wyner, A. (2010) 'IV Workshop on Legal Ontologies and Artificial Intelligence Techniques', *Towards Annotating and Extracting Textual Legal Case Elements*, 9 (http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.232.1531&rep=rep1&type=pdf#page=9)

Landes, W. and Posner, R. (2020). *Legal Precedent: A Theoretical and Empirical Analysis*. [online] NBER. Available at: https://www.nber.org/papers/w0146 [Accessed 21 Jan. 2020].

Javaid Nabi (2019). *Machine Learning — Probability & Statistics*. [online] Medium. Available at: https://towardsdatascience.com/machine-learning-probability-statistics-f830f8c09326 [Accessed 26 Jan. 2020].

7.Bibliography

Steven Bird, E. (2020). *Natural Language Processing with Python*. [online] Shop.oreilly.com. Available at: http://shop.oreilly.com/product/9780596516499.do [Accessed 26 Jan. 2020].

Anon, (2020). [online] Available at: http://outsourcingsurvival.com/tag/enterprise-information-archiving-eia/ [Accessed 26 Jan. 2020].

Citizens' Oversight Maryland---Maryland Progressives. (2020). *Category: Third-wayneo-liberals*. [online] Available at: https://www.citizensoversightmaryland.com/blog/category/third-wayneo-liberals [Accessed 26 Jan. 2020].

Brüninghaus, S., Brüninghaus, B. and Ashley, K. (n.d.). *Predicting Outcomes of Case-based Legal Arguments*. [online] Available at: http://www.lrdc.pitt.edu/ashley/bruninghaus.pdf [Accessed 27 Aug. 2019].

Disha Misal (2019). *How Artificial Intelligence Can Predict Trial Outcomes & Help Justice Prevail*. [online] Analytics India Magazine. Available at: https://analyticsindiamag.com/how-artificial-intelligence-can-predict-trial-outcomes-help-justice-prevail/ [Accessed 26 Jan. 2020].

Giuliano Giacaglia (2019). *How Law is being transformed by Machine Learning*. [online] Medium. Available at: https://medium.com/@giacaglia/how-law-is-being-transformed-by-machine-learning-3642045a0d34 [Accessed 26 Jan. 2020].

https://www.facebook.com/jason.brownlee.39 (2019). *A Tour of Machine Learning Algorithms*. [online] Machine Learning Mastery. Available at: https://machinelearningmastery.com/a-tour-of-machine-learning-algorithms/.

Itai Gurari (2017). *The Statistics of Winning (Motions in Litigation)*. [online] Medium. Available at: https://blog.judicata.com/the-statistics-of-winning-motions-in-litigation-5c4b60f126c5 [Accessed 26 Jan. 2020].

Itai Gurari (2018). *From Judging Lawyers to Predicting Outcomes*. [online] Medium. Available at: https://blog.judicata.com/from-judging-lawyers-to-predicting-outcomes-f46aedeb8684 [Accessed 26 Jan. 2020].

Lionbridge AI. (2019). *10 Best Legal Datasets for Machine Learning | Lionbridge AI*. [online] Available at: https://lionbridge.ai/datasets/10-best-legal-datasets-for-machine-learning/ [Accessed 26 Jan. 2020].

Netguru.com. (2018). *How Machine Learning Disrupts the Legal Environment*. [online] Available at: https://www.netguru.com/blog/how-machine-learning-disrupts-the-legal-environment [Accessed 26 Jan. 2020].

Ngo, M. (2017). *Classification of Legal Verdicts Using a Corpus Linguistic-Based Approach*. [online] Available at: http://arno.uvt.nl/show.cgi?fid=142434 [Accessed 26 Jan. 2020].

Premonition (2018). *The Future of Law is Digital | Artificial Intelligence in Law | Premonition Analytics*. [online] Premonition. Available at: https://premonition.ai/future-law-digital/ [Accessed 26 Jan. 2020].

Priyanka Bhilare, Parab, N., Soni, N. and Thakur, B. (2008). Predicting Outcome of Judicial Cases and Analysis using Machine Learning. *International Research Journal of Engineering and Technology*, [online] 9001. Available at: https://irjet.net/archives/V6/i3/IRJET-V6I362.pdf.

Stewart, M. (2019). *The Most Important Supreme Court Decision For Data Science and Machine Learning*. [online] Medium. Available at: https://towardsdatascience.com/the-most-important-supreme-court-decision-for-data-science-and-machine-learning-44cfc1c1bcaf [Accessed 26 Jan. 2020].

8.Appendix

8.1 Appendix A (Sequence Diagrams)

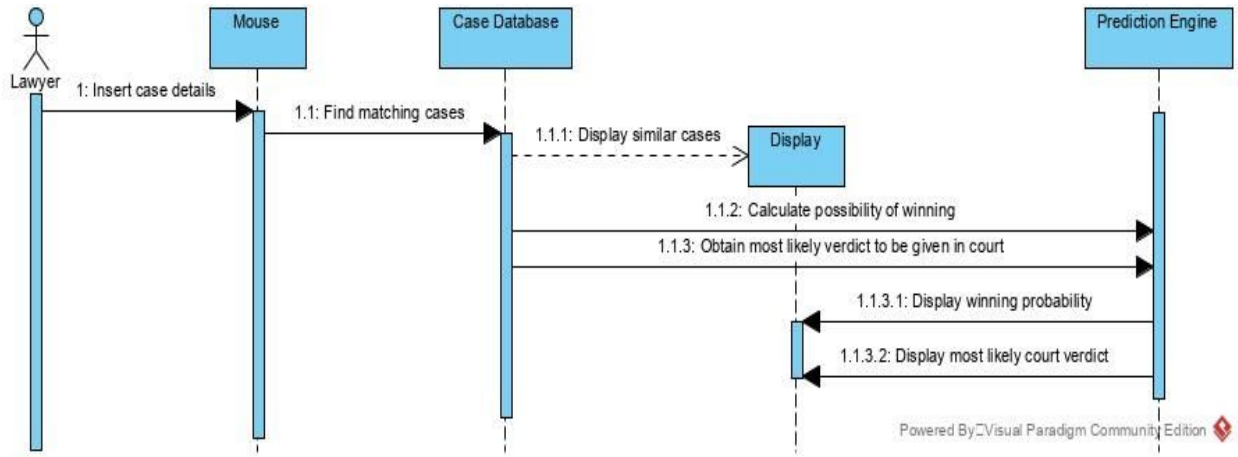


Figure 8.1. Sequence Diagram - Entire System

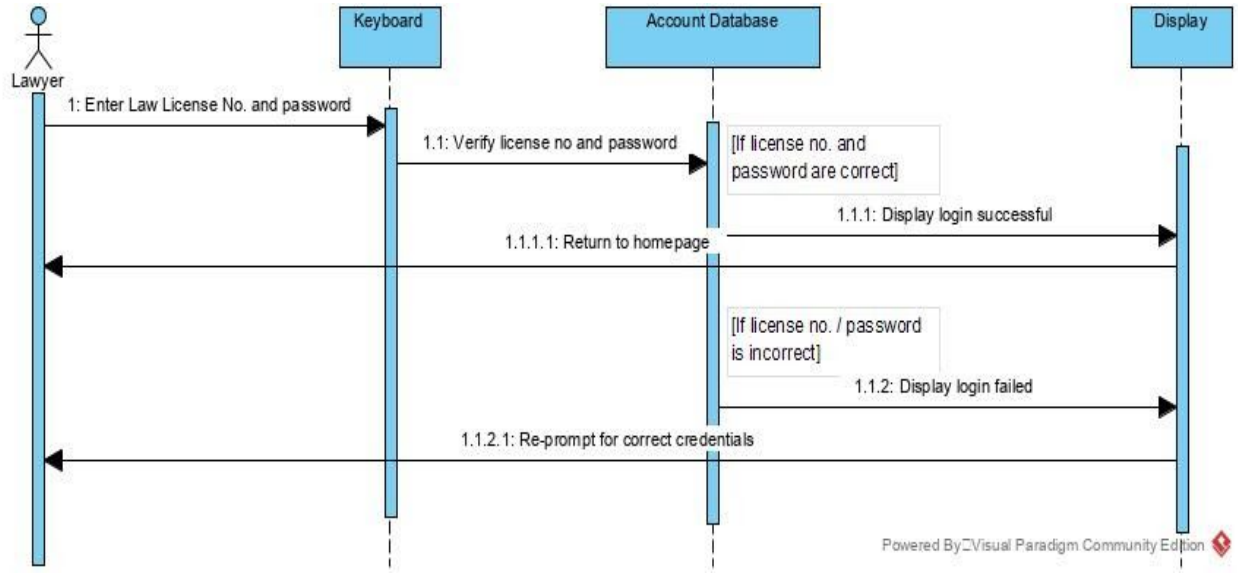


Figure 8.2. Sequence Diagram - Login Lawyer

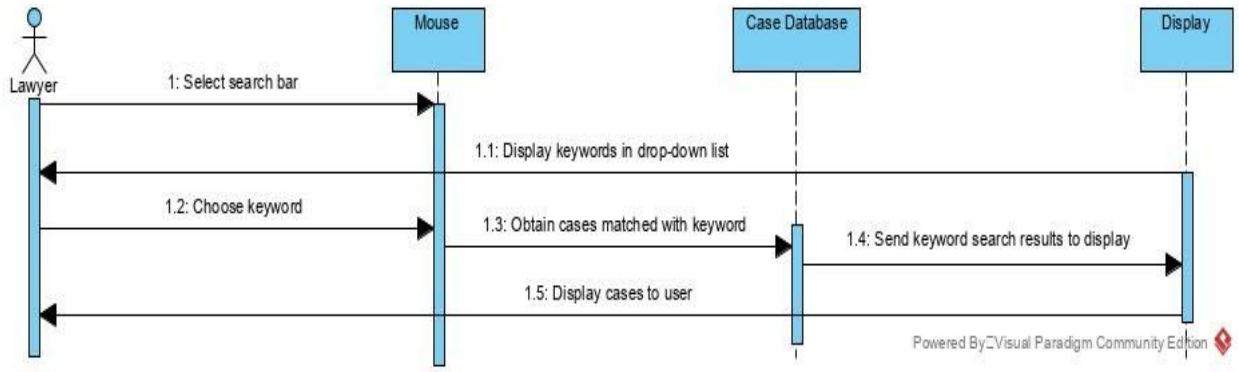


Figure 8.3. Sequence Diagram - Search by Keyword.

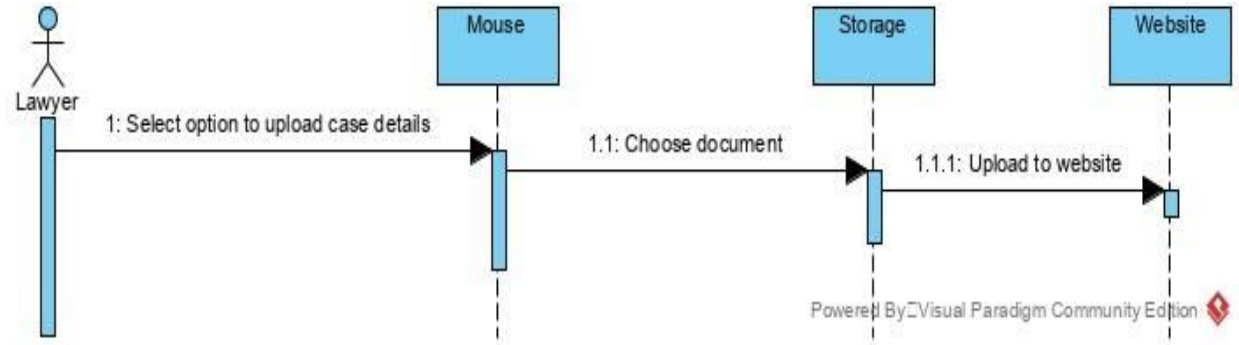


Figure 8.4. Sequence Diagram - Uploading a Case File

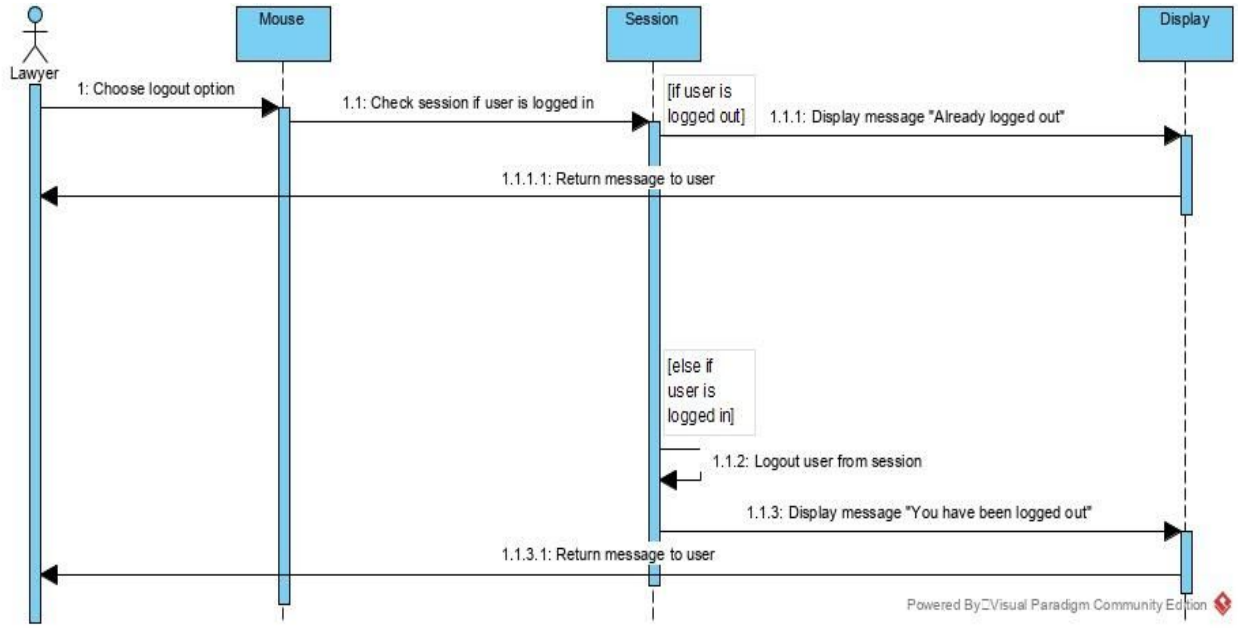


Figure 8.5. Sequence Diagram - Logout from System

8.2 Appendix B (Activity Diagram)

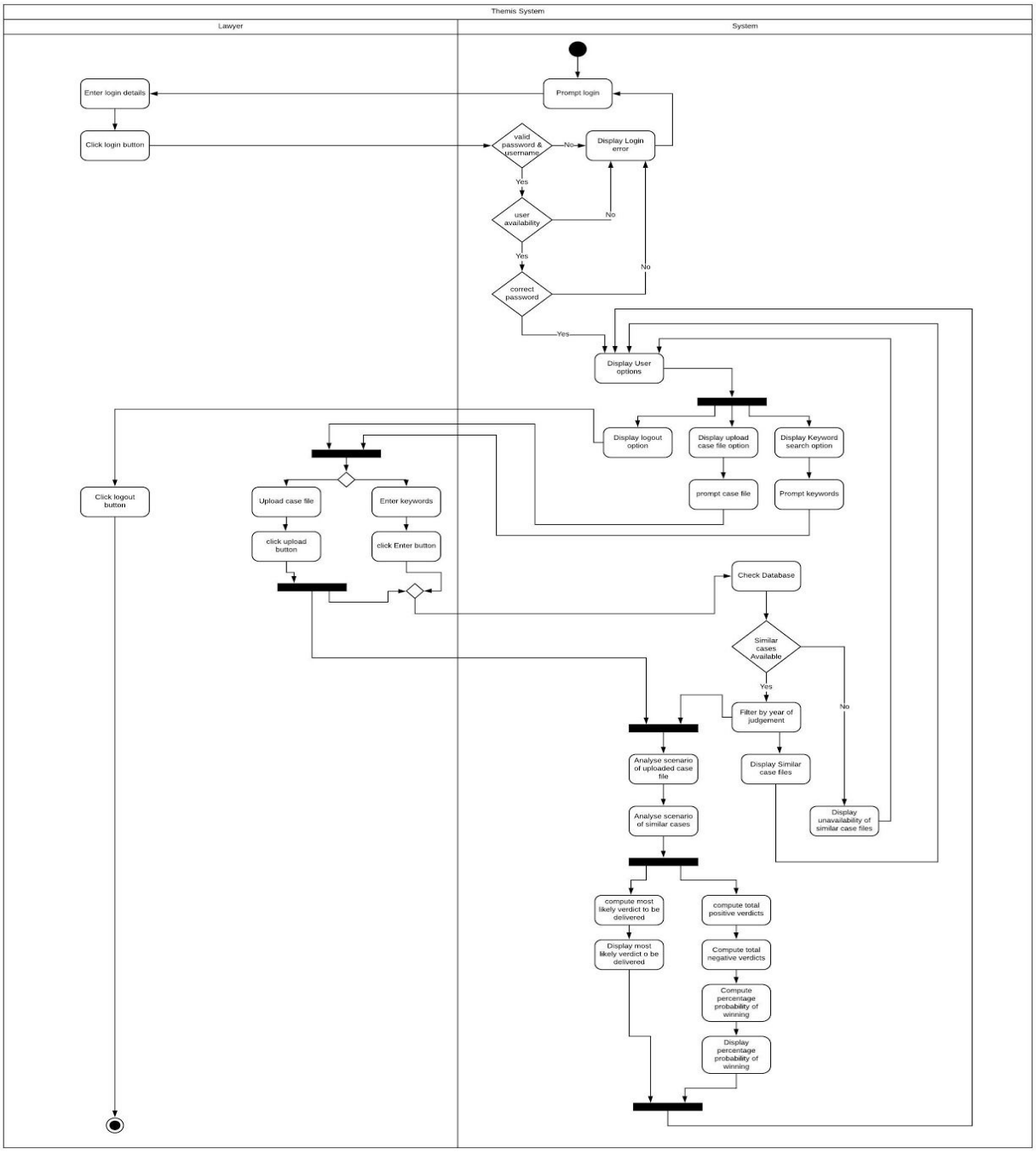


Figure 8.6. Activity Diagram

8.3 Appendix C (Work Breakdown Structure)

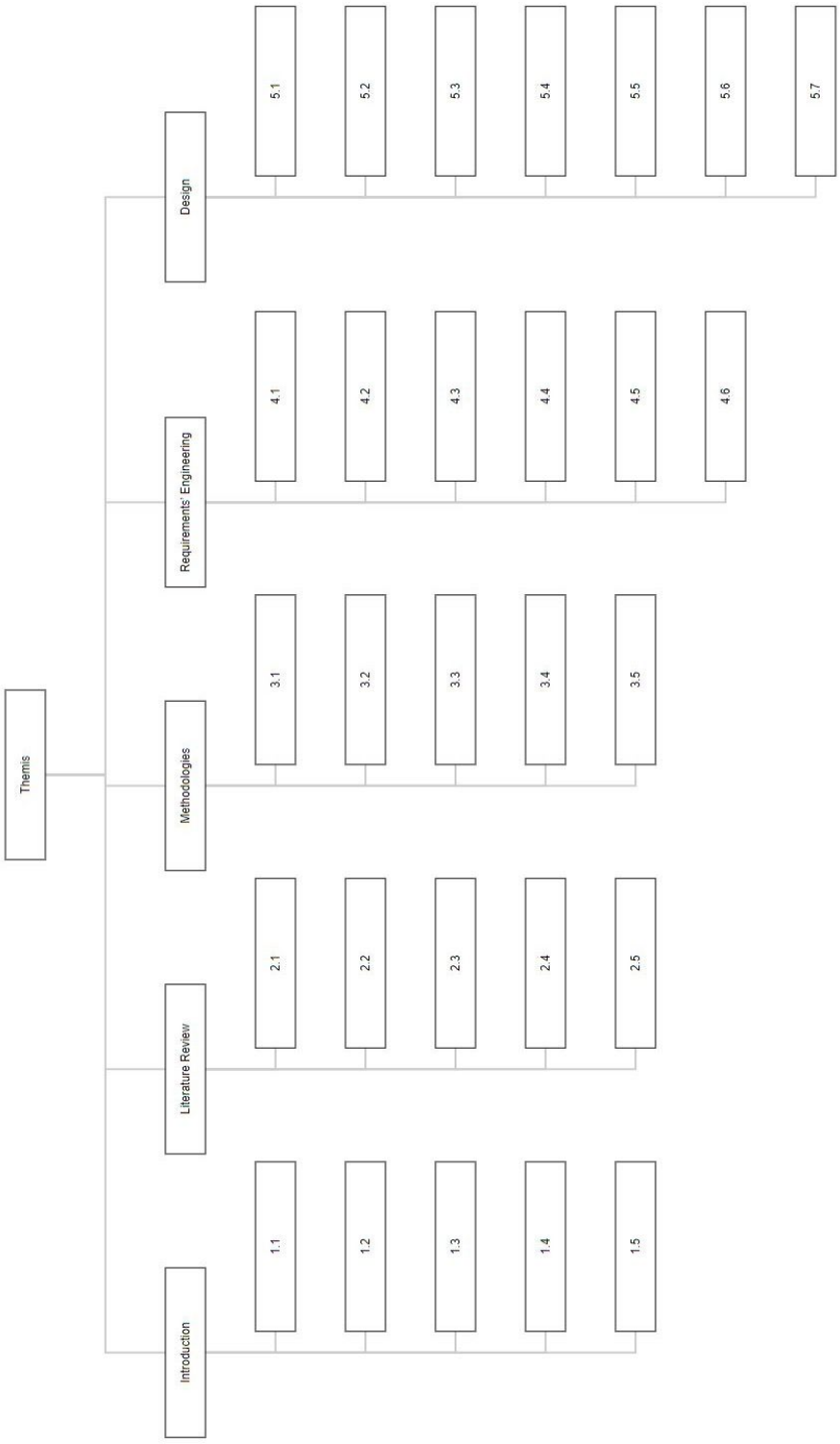


Figure 8.7. Work Breakdown Structure