MODELLING AND DESIGNING AN OPTIMUM RECOMMENDER SYSTEM FOR ONLINE MENTORING SERVICES

DR MUDIARASAN KUPPUSAMY AND VIJAYA SHREE RAJA SEKARAN

Submitted to Future Laboratory Sdn Bhd

ASIA PACIFIC UNIVERSITY OF TECHNOLOGY & INNOVATION (APU)

DECEMBER 2019

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ABSTRACT

The online mentoring platform allows potential mentors to connect with students, professionals, and entrepreneurs. The service platform for online mentoring is not equipped with features that suggest appropriate mentors. The research covers the aspects of online mentoring services, advantages, and disadvantages of the information behaviour models. These models aid in detecting the dynamics and behaviour of the customers. Thebehaviour and other features are used in the predictive analysis to suggest the mentors. Moreover, the research on millennials online behaviour will help outline their digital footprints. The online recommender systems in other platforms will be considered in modelling the algorithm for the smart matching feature on the test website. Thus, the research aims (i) To identify and build the customer behavioural model and use it in modelling recommender algorithm at a practical level, which has the subsequent impact on customer satisfaction, which is considered the business's success factor. (ii) To categorise the application development area focusing on recommender engine and user experience.

SECTION 1

INTRODUCTION

A digital mentoring service is a platform that connects the mentors and mentees through supportive and structured services by which the mentees are guided to develop the skills and achieve their goals. Online mentoring is becoming a popular hobby among tech-savvies and professionals. Beyond hobby, mentors volunteer in the mentoring activities, enabling them to spread their knowledge and experiences (FutureLab, 2019). Most digital mentoring has a professional volunteering model that connects with fellow mentees through the platform. The technology and internet-driven platform for online mentoring are typically webs and mobileapplications available for the mentees to use the services that are partly free of cost.

The journey of mentee and mentor: In the application platform, the mentee starts to provide necessary details to register themselves, create their profile in the application, and find suitable mentors based on their requirements, at this point where the journey of the mentee begins in the online mentoring platform whereas the mentors' journey is slightly different. Unlikethe mentee profile, mentors' profiles will have the profile validation check made by the platform service to approve and gauge the authenticity of the mentors (FutureLab, 2019). This platform also provides the avenue for professionals to establish a professional network. By networking, theyalso get endorsements, job opportunities and acknowledgements from their network, which is an added credential to their digital profile in this technology era (Clevers, 2018).

This research project intends to study the dynamics and user behaviour of the online mentoring platform. This platform's customers (users) are students, working professing and young entrepreneurs (FutureLab, 2019). The mentors are the experts in their field of interest that mentees prefer to have guidance. The companies that offer this platform can increase their customer base by understanding their behaviour and expectations. An empirical case studyon *FutureLab*, one of the companies in Malaysia that provides a digital mentoring service, will becarried out in this research. The behaviour of its customers and their feedback will be considered in building a suitable machine learning model of a recommender system which is the crucial feature related to business improvement, optimisation and customer satisfaction.

1.1 RESEARCH BACKGROUND

The online mentoring platforms have several features and workflows that enable users to perform the necessary activities to render the service optimally. The customers' behaviour differs from one platform to another, especially on the internet. The need for information, communication and the steps taken by the users to search the info is emphasised in many models (Leckie, Pettigrew, & Sylvain, 1996). The behaviour of the users canbe understood from the communication layer that consists of human-computer interaction (Wilson,1999).

1.1.1 THE NEED FOR ONLINE MENTORING

The mentee reaches out to the mentor for guidance which helps them accomplish and prove their uniqueness and talent. All the professional tasks, even the smallest of the tasks, need accreditation and acknowledgement to be considered a quality contribution. In many instances, the mentors provide the acknowledgement and endorsement of the tasks that are then considered to be completed successfully. Often the mentors are expected tofind the qualities and aspirations of the mentees. Although the mentees are less focusedduring the starting stages, mentors help them build the required skills (Clevers, 2018).

Thus, the need for a mentor and their impact on the deliverables are essential in the mentee's professional growth. The mentoring sessions can be carried out in many ways. Due to the advancement in technology and the internet, online mentoring service is becoming popular among netizens, especially millennials. The advances have helped beyond the geographical boundaries and provided a consistent association with the mentor. Moreover, theonline communities empower the mentees to find the appropriate mentors that match their profileand interests (Gottlieb et al., 2017).

1.1.2 POPULARITY OF ONLINE MENTORING

The need for online mentoring services has made the service popular among millennials. As the millennials are offered a wide variety of jobs and career options, they look for mentors with experience in the field who can guide them to make an informed choice (Brett, 2018). Mentors are digital leaders and experts with experience in the domain. According to the career success pyramid, they offer guidance to the mentee, which is one of the critical factors of career success.

As shown in figure 1, the career success pyramid emphasises that choosing a mentor is an important part of the professional journey. The mentors are generally the influencers and can grow your interest in learning, networking and building the skills (Brett, 2018). The users, such as students, aspiring entrepreneurs, and working professionals, look for web-based mentoring services as it helps them to pick their mentors based on their needs (Gottlieb et al., 2017).

Figure 1 Career Success Pyramid (Brett, 2018)

The advantages of online mentoring are also an important factor in its popularity. The pros of online mentoring such as connection beyond boundaries, sharing the information that can be accessed anytime, organising the meeting based on the time convenience of the mentors and mentees, are also popular (Kumar & Johnson, 2017). The differences in time-zone issues in connecting and understanding are the cons of online mentoring. The workaround for the cons can be implemented by appropriate planning, communication, and arrangements (Pillon & Osmun, 2013). Figure 2 summarises the pros and cons of online mentoring:

Pros of Online Mentoring	**Cons of Online Mentoring**
• Meeting time is Convenient • Information can be accessed anytime • No geographical barriers	• Time-zones may be different • Understaning takes long time • Connectivity issues

Figure 2 Pros and cons of online mentoring

1.1.3 EXPECTATIONS AND BEHAVIOUR OF NETIZENS

The netizens acquire knowledge of products, services, and other topics through the internet. They show a keen interest in using the internet for information search activities (Utkarsh,Sangwan, & Agarwal, 2018). The spread and access of information are made easy and fast due to the enhancements in the technology (MacKinnon, 2012). The online behaviour of the netizens, like commenting and reviewing the product and services, has a great impact such that they utilise good and bad reviews in the decision-making process (Prendergast, Paliwal, & Chan, 2018).

The online comments from the netizens have created a new objective for the sellers and marketers that is to maintain and improve their product and services, as the online reviews directly affect their brand value and reputation (Utkarsh et al., 2018). The netizens tend to get the knowledge, opinion and perception of the products and services from the others opinion and perceptions via their reviews and comments (Dörnyei & Gyulavári, 2016). Netizens trust the reviews and comments of the stranger and claim that they are genuine and cognitive in providing the reviews and comments despite their behaviour and personality (Li & Guan, 2013).

1.1.4 THE PURPOSE OF RECOMMENDER SYSTEMS

The recommender systems are often used as a sophisticated feature in search engine optimisation and other information search processes. The PEW Internet and American life project carried out an empirical study on internet usage. It is found in the study that the information search is the top activity, as shown in table 1. The internet has moved the traditional data search in the library to online platforms where the college students use the web indexes to effectively obtain the data from the web repositories (Reddy, Krishnamurthy, & Asundi, 2018).

Table 1 Usage of the Internet (Source: PEW)

Internet activity	Percentage
Search engine to find information	87%
Look for health-related information	83%
Look for hobby or interest research	83%
before purchase	78%
Buy a product	66%
Buy or make travel reservations	66%
Use online Ads or sites	53%

The recommender systems provide suggestions to the users that search for information on the internet. It is used in almost all static and dynamic web pages and applications as it allows for exact or close to the precise search terms of the users. This feature helps to fulfilthe expectation of the users as they look for appropriate information concerning their search. The recommender system is also one of the features that provide customer satisfaction, especially in SaaS (Software as a Service) domain. The customers are self-guided in learning the software product and its uses as the search engine and recommender models make the information retrieval.

1.1.5 EXISTING MODEL OF RECOMMENDER SYSTEM

Google's optimisation technique is the existing and popular optimising model, using recommender systems. When the user searches in the google search bar, the web crawlers duplicate the webpage and download those pages and associated link pages which is then brokendown to obtain any URLs (Uniform resource locators) and hyperlinks (Egri & Bayrak, 2014). Allthis information is kept in the server called store serve and the archive and word identification numbers (ID). These are used in sorting and accessing web pages. It also follows the page ranking algorithm to rank the web pages before the suggestion (Solihin, 2013).

The recent search engine optimisation technique uses a machine learning technology filtering algorithm to use the recommender system. In online businesses, machine learning models are important because of their ability to independently handle the data, learn from the previous computation process and provide reliable results that can be used for the business decision making process. There are many types of machine learning models, such as collaborative filtering, content-based filtering, classification models, and so on (Bobadilla, Ortega, Hernando, & Gutiérrez, 2013).

The factors considered in choosing the suitable algorithm are data model and structure, database and other integration technology, benchmark quality, and many other factors. The hyperparameters can define and build the finest recommendation model based on the business requirement. The correct combination of the elements and models will result in the best recommender system in supervised and unsupervised machine learning models (Bobadillaet al., 2013).

1.2 PROBLEM STATEMENT

Customers expect several factors from the mentoring service, especially in online mentoring services like mentor profile, meeting schedules, feedback, workarounds, etc. are to be matched with the customer profile before suggesting the mentors (Gottlieb et al., 2017). The existing networking applications have a profile match based on other factors that are not always similar to the mentoring services (Bobadilla et al., 2013). The mentoring services need a recommender system that reflects the customer dynamics and online behaviour.

The crucial success factor in any business is customer satisfaction. In the empirical case study of FutureLab, the test site provides mentoring services to mentees with a basic search engine model that suggests vaguely relevant results. The customers' search intentions and satisfactionare affected due to this model (FutureLab, 2019). The customer experience in the information search is also limited since they have not adopted the machine learning model. The existing model uses data from the same schema stored by the customer profile data. Therefore, data like customer rating, which is one of the important factors in the suggestion of mentors, is not used in the recommender model.

The reviews on the mentoring sessions by the customers are collected and stored in the unlinked table in the database, which is not used in any of the business or workflow changes. Likewise, the customer ratings are also collected and stored without any framework and business objectives as they are not used elsewhere. This approach gives unsorted recommendations, leading to a poor customer experience where the customers manually choose the mentors and schedule the meeting with the mentor (FutureLab, 2019).

1.3 RESEARCH QUESTIONS

The research questions are as follows:

1. What are the online behaviour and information search behaviour of the netizens?
2. What is the importance of digital mentorship among millennials?
3. What is a suitable machine learning recommender system?
4. What are the new features that can be added to the recommender system?
5. What technical workflow changes be implemented in our case study?

1.4 AIM & OBJECTIVE

The research aims to comprehend the dynamics and online behaviour of the customers in the mentoring platforms, which can be included in building a machine learning model of the recommender system. The customers' behaviour differs in the online platforms and when it comes to the online mentoring platform, it tends to have additional nuances like networking, endorsements, and more. Therefore, this research aims to derive the behavioural model specifically for the online mentoring platform, which enables the business and customers to have uninterrupted and accurate service.

The main objectives of the research are:

1. To develop the behavioural model of online mentoring service users
2. To model the parameters of an optimum recommender system for online mentoring service
3. To design and build the optimum recommender system for online mentoring service
4. To develop a rating model that can be integrated with the new recommender system for online mentoring service

1.5 SCOPE OF THE RESEARCH

The users of online mentoring platforms are aspiring entrepreneurs, students and working professionals. Since Malaysia offers a wide range of career opportunities, users are presented with multiple career choices (Kaur & Kaur, 2008). The fresh graduates and theprofessionals willing to change their work domain look for mentors to guide them in the proper direction to achieve their goals. Thus, FutureLab – a startup in Malaysia, is identified asthe most suitable online mentoring platform. It offers a digital mentoring service and a portal connecting potential millennials and experts. It uses technology and aims to reduce its users' knowledge and skill gap.

An empirical case study is carried out to build the behavioural and recommender model for its digital mentoring platform. After understanding the customer dynamics, the behavioural model is derived. The university students test the existing model, and the feedback and thebehavioural model is used in building the new recommender model. The software tools and languages used in this research are shown in Table 2.

Table 2 Software Tools & Languages

PURPOSE	TOOLS / LANGUAGES
Project Management	: Trello
Database View	: PGADMIN4
Data Retrieval from database	: Postgres SQL
Data Exploration & Story	: SAS Studio & Tableau
Integrated Development Environment	: Anaconda Jupyter notebook
Data Preparation	: Python3
Recommender system ML model	: Sklearn (Python3)
Visualization	: Tableau

1.6 SIGNIFICANCE OF THE RESEARCH

The evolution in the job market opportunities has led the users to seek guidance from experts in the domain. The hiring managers and employers expect that the candidate should possesstechnical and practical skills and academic skills (Kaur & Kaur, 2008). Thus, the students look for the guidance of the mentors who are the experts in their domain to help them understand the employer's skill expectations.

The entrepreneurs look for guidance in their technological and business models to implement their idea and build a successful enterprise. They believe in the peer reviews and suggestions which impact their startup business performance (Chatterji, Delecourt, Sharique Has, & Koning, 2018). The working professionals look for mentors to upskill their profiles or change their domain.When there is less job satisfaction for the working professional, they tend to improve their skills to perform better. They might also look for changing their domain based on their recent experiences. The mentor helps them make the crucial choice of changing their domain (Valaei& Rezaei, 2016).

These varied expectations of the users can be provided in the digital mentoring platform services. A feature like recommender systems has a huge impact on customer satisfaction where it provides a precise match of the mentors to the users. Also, the recommender system parameters reflect customer dynamics and behaviours. One of the features is the rating of the mentoring session that can be effectively used in suggesting the mentors to the users. Thus, the recommender system collectively enhances services and customer satisfaction.

SECTION 2
LITERATURE REVIEW

This chapter consists of a review of existing concepts in the behavioural study, empirical study and the research gap in online mentoring, customer behaviour and recommender systems. These are the area in which the empirical case study will be carried out. This review intends to understand the existing research and models in the mentioned areas, which will aid in building and implementing the recommender model and the rating model for the test site.

2.1 BEHAVIOURAL STUDY

The customers of the online mentoring platforms are considered while choosing the models used in this behavioural study. The customers intention, need, and behaviour of information search are carefully studied.

2.1.1 INFORMATION BEHAVIOURAL MODELS

Behavioural models are used in understanding the complexities of information retrieval, understanding and communicating the retrieved information (Robson & Robinson, 2013a). The different types of behavioural information models are discussed in this section.

2.1.1.1 ELLIS'S MODEL

The empirical analysis of the survey data from the domains of education private and government enterprises was carried out by Dr Ellis in devising this framework. The data was collected from the professionals of the mentioned domains (Ellis & Haugan, 1997). These professionals seek information by following the set of activities in any order. The activities are choosing the topic, carrying out the research on the topic, segmenting the data based on the quality, monitoring the changes made to the topic and finally obtaining the required information. (Ellis & Haugan, 1997).

The stage process version of Ellis's model, as shown in figure 3, consists of six levels. The information is monitored and browsed at the same level, after which it is differentiated. Once the information is spotted, it is extracted and checked for correctness which marks the end of the staging process. The disadvantage of Ellis's model is that the criteria such as information source and how it is communicated across the team are not covered.

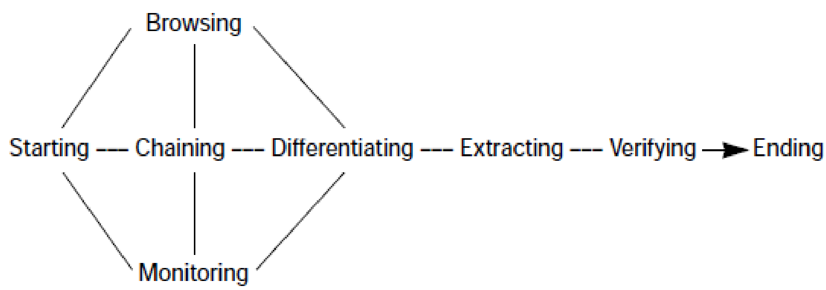

Figure 3 Ellis's Model (Wilson, 1999)

2.1.1.2 WILSON'S MODEL

There are different models designed by Dr Wilson, who is primarily subjective to the interactionof information by the users. The information search activity is carried out with cognition, which aids in understanding the information and knowledge discovery, as shown in figure 4. The information source can be either formal or informal. A formal source of information consists of a library and online accredited works. The informal source consists of an internet source and peer communication.

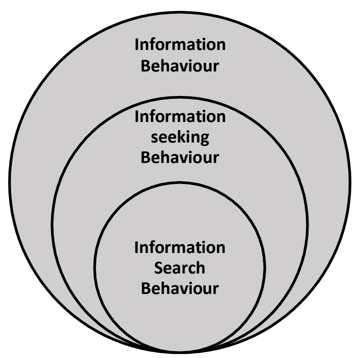

Figure 4 Wilson's model (Wilson, 1999)

There are three models, and the final model is shown in figure 4, which has three nested levels:

(i) Information behaviour: This level consists of general human behaviour in the information search and human-computer interactions
(ii) Information Seeking behaviour: This level consists of methods of seeking information
(iii) Information search behaviour: This level consists of information interaction and communication

The levels can be utilised in building the multi-disciplinary contents as the model is flexible and inclusive of all the levels (Tury, Robinson & Bawden, 2015). The subject centre has to be diligently selected as it reflects the objectives of the information search model.

2.1.1.3 INGWERSEN AND JARVELIN'S MODEL

There are seven levels in this model. Each level consists of the factors that affect information search behaviour. The user mentality, peer influence, context from culture, social and organisation, the need for information and the technology that provides this information are the factors of this model(Ingwersen & Jarvelin, 2005). As shown in figure 5, this model incorporates cognitionin seeking and providing information. The disadvantage of the model is that it is completely dependent on human cognition and viewpoint. Hence, the retrieved information might convey a different meaning than the actual meaning and depends on the user (Robson & Robinson, 2013).

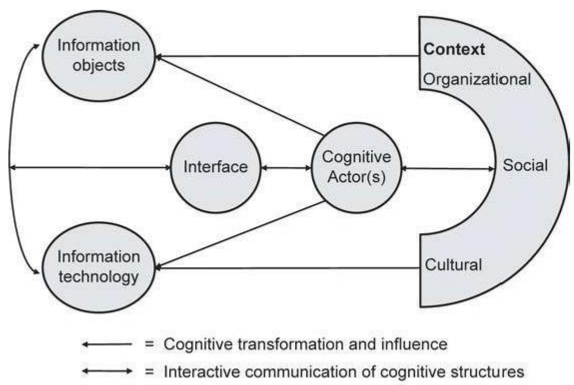

Figure 5 Ingwersen and Jarvelin's Model (Wilson, 1999)

2.1.1.4 LECKIE'S MODEL

There are six levels in this model and the defined communication flow. The levels in Leckie's model are the work roles, tasks, characteristics of information need, source of information, awareness of information, and the outcomes. These levels are followed in the information retrieval process by the experts and scholars in their domain. Mostly, the information sought does not match the user's requirement due to the improper availability of information as the model is built fora specific domain (Leckie et al., 1996). Maintaining the health of the domain data is a separate workflow by itself. Since the retrieved information doesn't match the requirement of the objective and the topic of data retrieval is modified. The model is depicted in figure 6, which has the defined flow of communication between the levels.

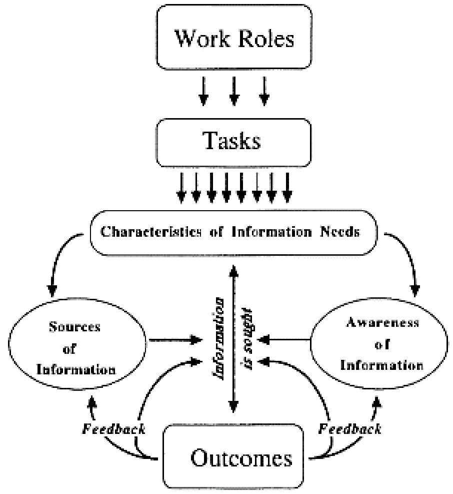

Figure 6 Leckie's model (Leckie et al., 1996)

The drawback of this model is that only the domain expert can successfully retrieve the information as only those users understand and are aware of the information (Leckie et al., 1996). Thus,the existing models of behavioural information search are studied, and the levels and features of the model are thoroughly examined.

2.1.2 ROLE OF TECHNOLOGY IN ONLINE MENTORING

The advancements in technology have increased the opportunity in all the aspects of the professional. It is the foundation of all the hardware and software accomplishments that empowers the tasks that we carry out in our professional life. Communication and information technology have created an opportunity for students to learn online. It mainly uses the online medium to connect the experts (mentors) and the mentees (Kumar & Johnson, 2017).

There are applications for online mentoring that offer end-to-end service in a single platform.Our empirical case study is carried out on one of such services based in Malaysia. FutureLab uses technology to integrate various activities in completing a single session through a digital mentoring service (FutureLab, 2019). The user's entire journey is defined, built, and integrated using various technologies.

2.1.3 MILLENNIALS KNOWLEDGE QUEST

The user's online behaviour is influenced by their experience and domain knowledge which will also be reflected in their information search for academic-related topics. The information is made readily available for scholars and researchers in various forms as the millennials choose online information over offline information (Utkarsh et al., 2018).

They prefer online materials because of the genuine ratings and commentsof their peers. Also, they do not trust the extreme positive and negative comments on social media (Prendergast et al., 2018). As the final learning outcome is found to be the same in online and offline training, the millennials prefer online courses due to the advantages such as affordable fees and flexible timings (Stack, 2015).

2.1.4 ONLINE MENTORSHIP - THE BUZZWORD

Online mentorship is the casual hobby of the professionals that guides the users intheir success path. The popularity and the advantages of this concept, as discussed in section 1.1.2, are the primary reasons for this buzzword - online mentorship. Internet technology has madeeasy access to the service, and the online communities provide an instant answer to most common queries. Thus, factors such as a professional hobby, ease of access and community support are collectively responsible for the popularity of online mentorship.

2.1.5 ADOPTION OF ONLINE MENTORSHIP AMONG NETIZENS

The bonus of an online mentoring service is that it can demand connection across boundaries which opens various opportunities to the users through the optimum utilisation of the platform (Wang, 2007). The platform does not fix the interaction among mentor and mentee, and it supports convenient session interaction based on the users. Once the session is completed, the mentor highlights the feedback and areas of improvement.

This concept is interesting as the mentor and mentee tend to solve real-world gaps and improve their knowledge and skills. The mentee's work can be done live to the mentor, which enables them to monitor the mentee performance and provide instant feedback and areas ofimprovement to the mentee (Martin & Bolliger, 2018). Again, the factors such as easily availableinternet and online services boost the adoption rate of the platform among netizens. Thus, online mentorship and service are widely adopted by netizens.

2.2 EMPIRICAL STUDY

The empirical study on modelling the recommender systems is carried out in this section. The data, features, and model parameters are the main focus of this empirical study.

2.2.1 THE IMPORTANCE OF MACHINE LEARNING MODELS

Machine learning helps the business run its operations without external instructions and optimise its profit and investments. The importance of machine learning is mainly due to its advantages, such as increasing the productivity, efficiency, speed, and accuracy of the process, improving data-driven decisions, and so on (Butner & Ho, 2019). As businesses strive to make their customers happy, machine learning can provide accurate predictions and easy fault detection so that the organisation can focus on the areas of improvement. Machine learning can beeasily integrated into any type of workflow.

2.2.2 MACHINE LEARNING MODELS IN RECOMMENDER SYSTEMS

Many machine learning models can be used in building the recommender model. The empirical case study will be discussed in this section.

(i) **Content-based recommender system:** This modelling is carried out on anime recommendation where the system learns through similarity library for the recommender system (Caeser, 2018).
Data: Anime recommendation database
Parameters: Name, Genre, rating

(ii) **Collaborative filter algorithm:** This is carried out on amazon.com's item- based collaborative filtering algorithm (Linden, Smith, & York, 2018). Since the calculation prescribes profoundly corresponded equal things, proposal quality is excellent. The calculation likewise performs well with restricted client information, delivering great proposals for the business.
Data: Amazon.com's dataset
Parameters: Item search, page traffic, user purchased and rated items

Among these two types of machine learning model, content-based recommender system will be more suitable for our case study as the parameters and library are similar.

2.2.3 TEST WEBSITE EXISTING MODEL AND USER INTERFACE

<u>EXISTING MODEL</u>

ACCOUNT TYPE:

MENTEE

HIGHLIGHTED – ML AREA

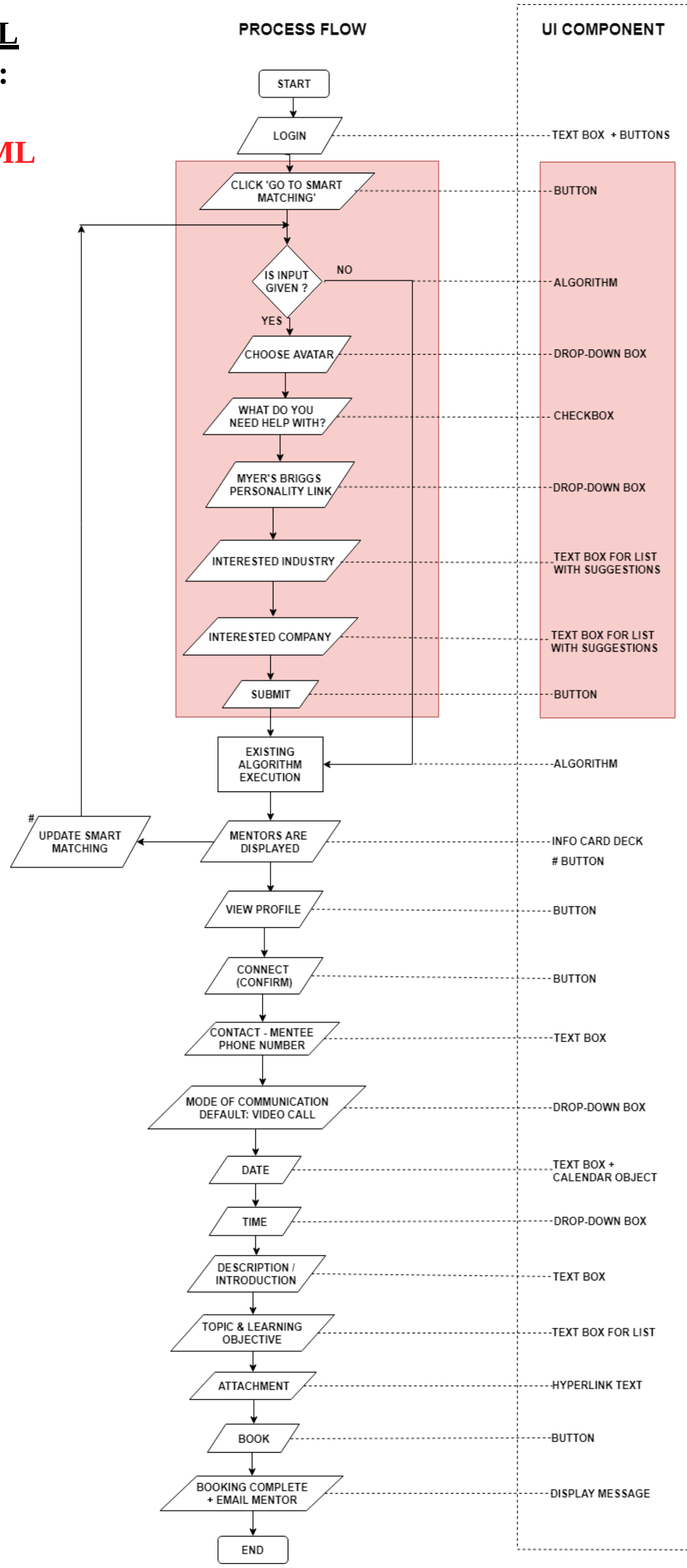

Figure 7 The flowchart of existing model in the test site

The flowchart of the existing test site is thoroughly analysed, and the potential area of machine learning is highlighted in figure 7. The digital copy of the flowchart is also provided on CD. The user interface of the test site is also analysed as it contributes and reflects the ratings of the customer experience and satisfaction.

2.2.4 RATINGS AS PERFORMANCE METRICS

Millennials have product knowledge and are capable of understanding its value. They possessprior product knowledge or begin their learning instantly over the internet (Awasthy, Banerjee, & Banerjee, 2016). It is observed that the product ratings and reviews are influential factors of the business which drives the business owners to maintain and improve the standard ofthe product and services (Dixit, Badgaiyan, & Khare, 2017).

The rating and reviewing of the product are considered to be the planned behaviour of the user where their activity is pre-planned. It expresses the appropriate emotions of the user who rate the features of online mentoring services. These features affectthe mentor's profile as the mentee rates the mentor and vice versa. It is an important factor that provokes the service providers to maintain the standard and improve it further.

2.3 RESEARCH GAP

The research gap in the behaviour models and the empirical models will be highlighted in this section. The research gap in the behaviour model is that the models are not built to fulfil the gap of the previous model. Among the four models, Leckie's model is the most appropriatemodel for our case study. But it is not exclusively defined for online data sources, and the cognitiveactors are supposed to be the expert in the domain. Also, the rating feature is not incorporated intothe information-sharing behaviour model.

The disadvantages of the summary tables are also the research gap of the model where only some of the features communication level and information source are fixed. The research gap in the empirical study is that the recommender model is not specifically built-in unsupervised learning techniques. It incorporates the theoretical parameters and is often not associated with business objectives and goals.

2.4 SUMMARY

The summary of the behavioural studies are tabulated as follows:

Table 3 Summary of Literature Review

SUMMARY OF BEHAVIOURAL MODELS				
Features	**Ellis's model**	**Wilson's model**	**Ingwersen and Jarvelin's model**	**Leckie's model**
Number of Levels	6	3	7	6
Level / Activity Names	1. Browsing 2. Chaining 3. Monitoring 4. Differentiating 5. Extracting 6. Verifying	1. Overall Behaviour 2. Information seeking behaviour 3. Information search behaviour	1. Information objects 2. Information technology 3. Interface 4. Cognitive actors 5. Organisational context 6. Social context 7. Cultural context	1. Work roles 2. Tasks 3. Information needs 4. Information source 5. Awareness of information 6. Outcomes
Is communication level included?	No	Yes	Yes	Yes
Advantage	+ Activities can be carried out in any order	+ Cognitive component + Defined information source + Defined communication in sub-layer	+ Cognitive component + Defined flow for communication	+ Defined information source + Defined flow for communication
Disadvantage	- No defined flow of communication - Undefined Information source	- Too general to build a specific model	- Difference arises between the conveyed and actual meaning of the information	- Only the domain expert can retrieve the information

2.5 PROPOSED ONLINE SEARCH BEHAVIOURAL MODEL

The enormous growth of information on the internet has enabled users to shift their information search medium from offline resources like libraries, journals, and newspapers to online resources like information from websites and e-libraries. The search behaviour model proposed in this study consists of an information source to be completed online. The information available online can be from trusted or non-trusted resources. The user holds the risk of exposure to non-trusted information resources.

The proposed model is shown in figure 8. The advantages and disadvantages of the existing behavioural models are analysed from the summary table 3 before building this model. This model depicts the information search behaviour, online user behaviour, and the new information-sharing behaviour level. One of the main factors of the reproduction of information in onlineapplications is the information-sharing behaviour of the users.

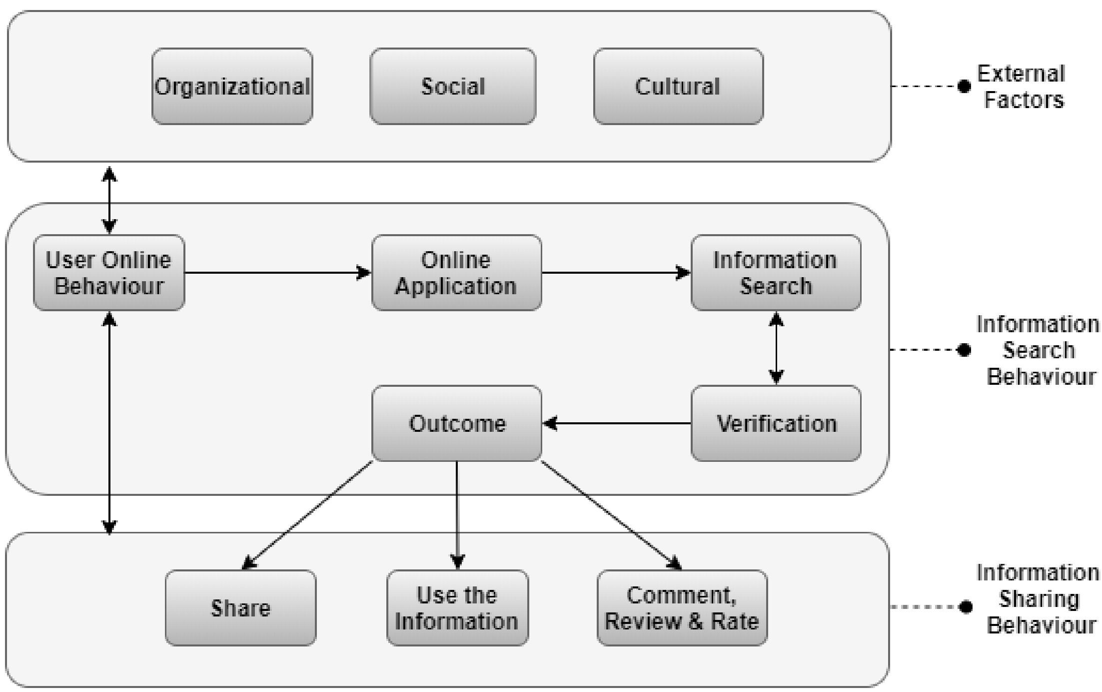

Figure 8 Proposed model for Online Search Behaviour

This model consists of three levels. The external factor level is emulated from Ingwersen and Jarvelin's model. The behavioural search level is from Wilson's model. The sharing behaviour is the new level added in the model.

2.5.1 DESCRIPTION OF THE BEHAVIOURAL MODEL

The proposed behavioural model for online search behaviour of the users consists of three levels:

(i) External factors – The user's need for information is influenced by organisational, socialand cultural factors. The user's role in each of these factors is the main reason for their information requirement.

(ii) Information search behaviour – The user will search for information in the online application with the help of a search engine that recommends the most relevant and appropriate information based on the search term. Once the information is obtained, the user verifies whether the information is useful or not. The outcome of verification is the next level.

(iii) Information sharing behaviour – The user will generate the new information by commenting, rating or reviewing the obtained information and sharing it back in the online medium itself.

2.5.2 ADVANTAGE OF THE MODEL

The advantage of the proposed model is that it provides the user behaviour, especially in the online information arena, which is not addressed in the existing behavioural models. The important user behaviour of online information sharing is added as a new level in this model. This sharing behaviour affects the factors of information search and, in turn, affects the online user behaviour. This level is one of the main reasons for (re)production of useful information for users such as comments, reviews and ratings, which enables the other users to make an informed decision.

2.5.3 DISADVANTAGE OF THE MODEL

The disadvantage of the model is that the user's exposure to information from non-trusted resources cannot be devised, and it affects their information-sharing behaviour. The recommenderengine plays a vital role in the online information search by providing appropriate information to the users. However, as the content generation is made open-source, which affects the credibility ofthe information, this increases the users' risk of exposure to non-trusted information beyond the control of online search engines.

SECTION 3

RESEARCH METHODOLOGY

This chapter consists of the data mining methodology used to research and develop themachine learning recommender system. The following sections also discuss the research approach, framework, data collection approach, data sampling, and research plan.

3.1 RESEARCH APPROACH AND FRAMEWORK

This research follows the quantitative framework design, mainly consisting of experimental design. It implies the exploration of data with evident perception rather than hypothesis. Frequently, this sort of research is carried out with the help of a dataset. The case study data is obtained from the business owner after performing research and framing objectives. It consists of large sample sizes and aims to find the optimum insight from the data. There are many types of quantitative research, such as surveys, experiments, etc.

The research objective to build the optimum recommendation model can be accomplished by adopting this framework, where the data is experimented with and modelled as per the customers' behaviour. The main tasks in quantitative research are observing, describing, and analysing the data used in the case study. Also, the mathematical modelling will be made out of the data, and hence the data transformation and preparation is the crucial steps that need to be followed. The outcome of this quantitative research is then integrated into the existing workflow of the controller in the test website, as shown in figure 9.

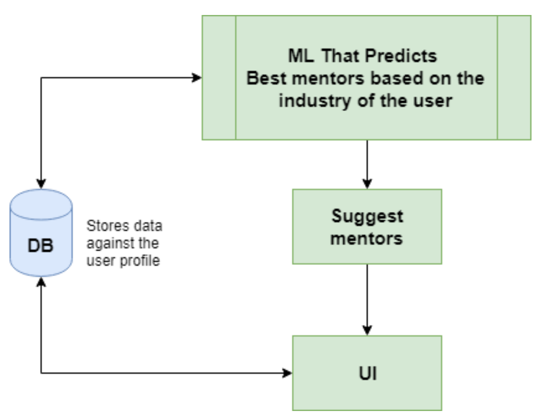

Figure 9 Integration of quantitative research and workflow

3.2 DATA COLLECTION APPROACH

The structured CSV data is collected from the FutureLab company, and thus, the secondary data collection is carried out in this research. The data collection plan and approach is shown in the below table 4:

Step 1: the researcher and the company sign NDA (Non-disclosure agreement).

Step 2: Understanding the business model and the data environment

Step 3: Researcher requests for the data and communicates to the company

Table 4 Data Collection Plan

Start date	Task	Task Details
9th March 2019	Business Data	Discussions on Data extraction and Visualisation
13th March 2019	Data Dump was seeded in pgadmin4	pgadmin4 was used to view the DB structure of futurelab.my
5th April 2019	New Data Dump was shared	FutureLab shared the new dump via Trello (Sprint). The dump was then uploaded to PGADMIN4 Database via Postgres SQL to analyse the tables
7th April 2019	Additional tables – data Dump was shared	Requested the data from required tables after reviewing the data provided on 5th April
10th April 2019	Study about FutureLab database	The analysis on the tables required for FrankBot was explored
18th April 2019	Additional tables – data Dump was shared	There were few tables missing in the dump shared on 7th April.
24th May 2019	Requested CSV format with table names	The mentioned attributes in Trello (Sprint) was requested in CSV format
28th May 2019	Reminder for the data	Raised the reminder in the Trello (Sprint) for the CSV file
	CSV file is shared	FutureLab shared the CSV file in Trello (Sprint)
3rd Jun 2019	Updated CSV file is shared	- Requested the CSV file with Mentee and Mentor ID - FutureLab shared the updated CSV file in Trello (Sprint) as requested
11th Jun 2019	Requirement finalisation	Data requirements of the model were finalised

3.3 SAMPLING

Once the data is obtained, it is explored in its file format view to understanding its structure,data type and schema design in the database. The sample of 2960 is shared in the dataset used to model the recommender system. Other common insights from the dataset, such as several sessions, ratings, mentor ID, mentee ID and their industries, can be known. The company primarily obtained the data from the direct registration of users on their website.The complete steps in data acquisition are explained in the methodology.

3.4 DATA ANALYSIS METHODOLOGY

This research follows the CRISP-DM methodology. CRISP DM refers to the Cross-industry standard process for data mining, which consists of a common process in data mining and modelling the data. The life cycle of the CRISP-DM has six stages, as shown in figure 10. It revolves around the data, and it is important to note that the life-cycle of CRISP happens inside the DM, represented using the outer cycle. Because of the continuous insights that data mining provides for business decision making and other development processes.

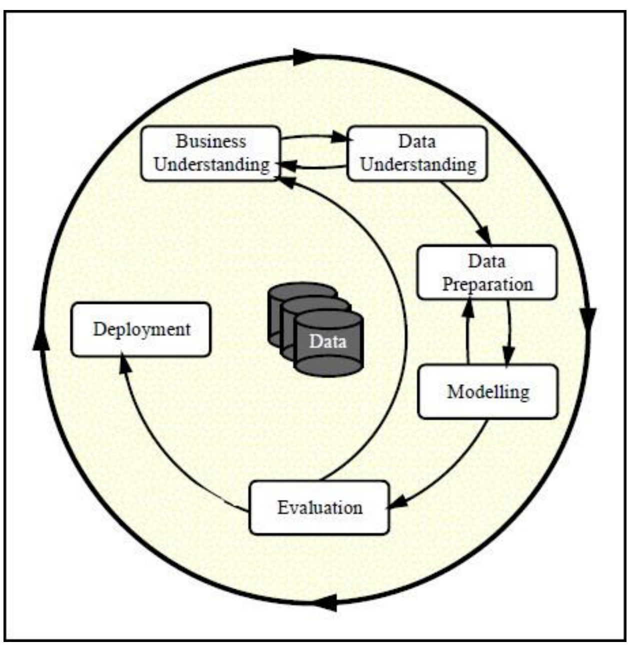

Figure 10 CRISP-DM Methodology

1) **Business Understanding:**

The business objectives are determined after understanding the business model of our case study company – FutureLab. The respective background study is also carried out where the business success criteria are customer satisfaction. The existing models in the test site are examined, and an improvement plan is devised as the research objectives. The project plan is made, which is discussed in section 3.5.

2) **Data Understanding:**

The secondary data collection is made as per the plan, as shown in table 4. Once the data is acquired in the *.dump* format is loaded to the pgadmin4 server. It is viewed in the pgadmin4 using Postgres SQL language for data retrieval. The data is explored from 178 tables by querying and viewing the data. Figure 11 shows the data exploration report.

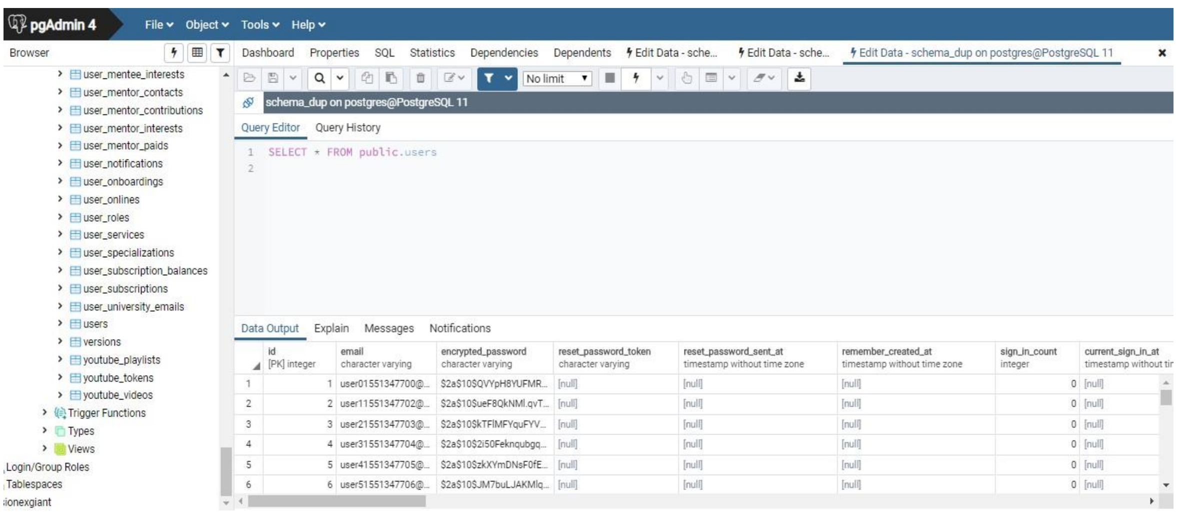

Figure 11 Data exploration report

3) **Data Preparation:**

The activities to build the dataset used in the modelling are pre-processed and prepared in the SAS Studio. The data preparation takes most of the analysis time, which includes the tasks such as data cleaning, missing value detection and imputation, outlier detection and handling, data transformation and other related tasks feature engineering and selection will be carried out to build the final dataset. The data chosen had the missing values in mentor_rating and mentee_rating values are imputed using the mean imputation method.

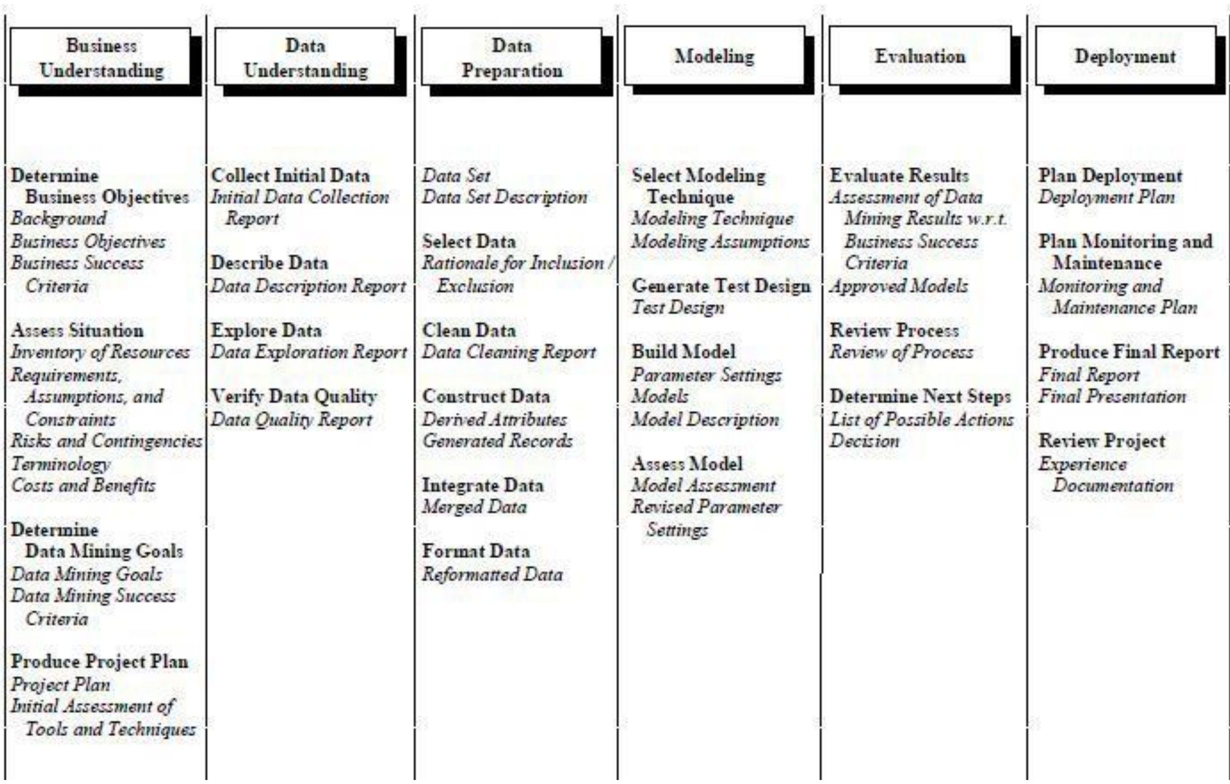

Figure 12 Overview of the tasks in CRISP-DM (Wirth, 2000)

4) Modelling:

As discussed in the literature review, the various modelling methods for the optimum recommender system will be experimented with in this step using the final dataset. The recommender model will be built in the integrated development environment called Anaconda Jupyter Notebook using the python3 language. The software compatibility is checked concerning the testing website's IT infrastructure.

5) Evaluation:

The data will be sampled into train and test datasets, where the modelling uses the training dataset, and this stage uses the test dataset. The model is also programmed from computing the accuracy. The evaluation stage can also be followed by the parameter tunning stage.

6) Deployment:

The developed model will then be integrated into the existing workflow of the test site. The website developer will incorporate the model in the website's controller.

3.5 RESEARCH PLAN

The planning is the first step in the project management, where the entire tasks involved in the research and modelling is included. This study is planned so that it is marked as a milestone for the important activities. The two types of planning are shown in figure 13, where one of the plans can be chosen based on the objective of the research (Wang & Gibson, 2017). In our research, plan A is chosen as we are intended to achieve the modelling result.

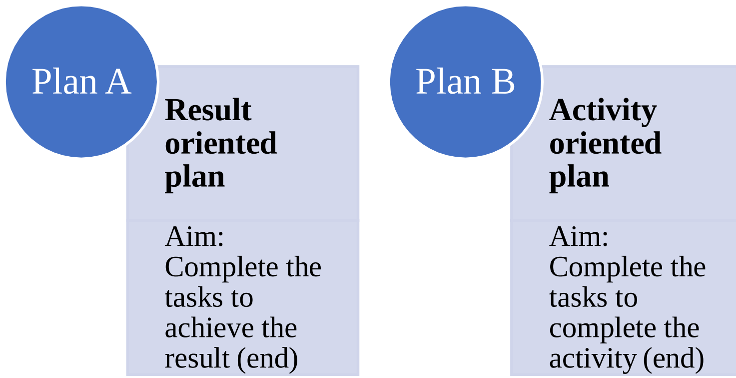

Figure 13 Types of research plan

There are 4 milestones in the research:

Table 5 Milestones

Milestone
(i) Gathering data – There are two subtasks: Gathering data from users & storing the information from the backend.
(ii) Data Analysis – Data analysis is the last step in the data mining process where the processed data is interpreted in identifying the features and outliers
(iii) Knowledge Base – This is the database maintained by the knowledge engineer. This should frequently be updated such that the machine exhibits the desired steps.
(iv) Training the model (Modelling) – Training, the model, is crucial in the machine learning the process as it moulds the model to perform with optimum results.

The 40 weeks project timeline created consists of the tasks and sub-tasks of the research. The trained model has been shared with the FutureLab team for integration and production as per the plan. The following figure 14 is the Gantt chart of this project:

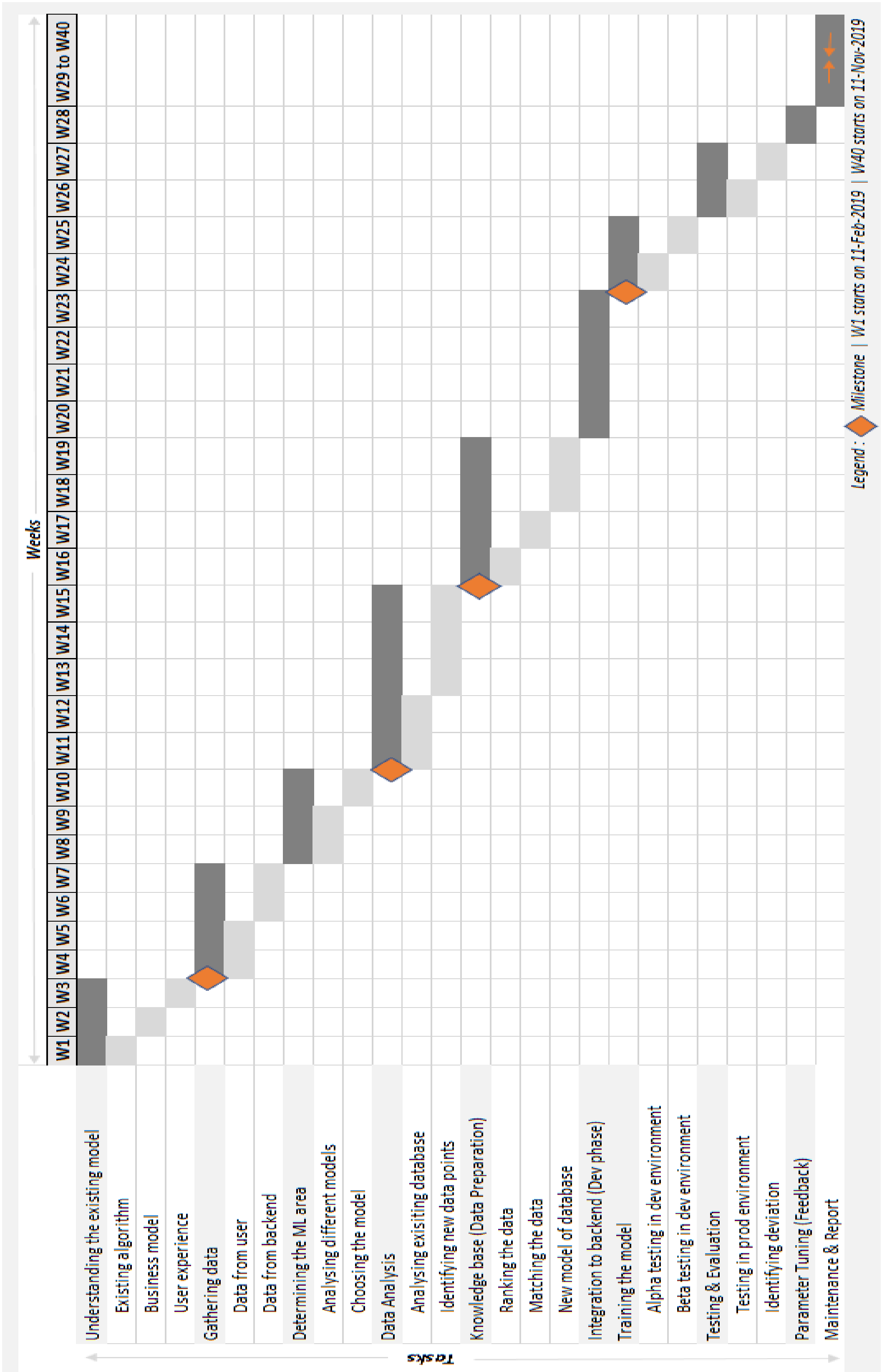

Figure 14 Gantt Chart - Research Plan

Project Management Dashboard- Tracking the milestones and other tasks in Trello:

The following figure 15 shows the interactive dashboard used as the sprint board in the project management software called Trello. The research and the development team promptly update the projects' tasks at FutureLab to track the progress.

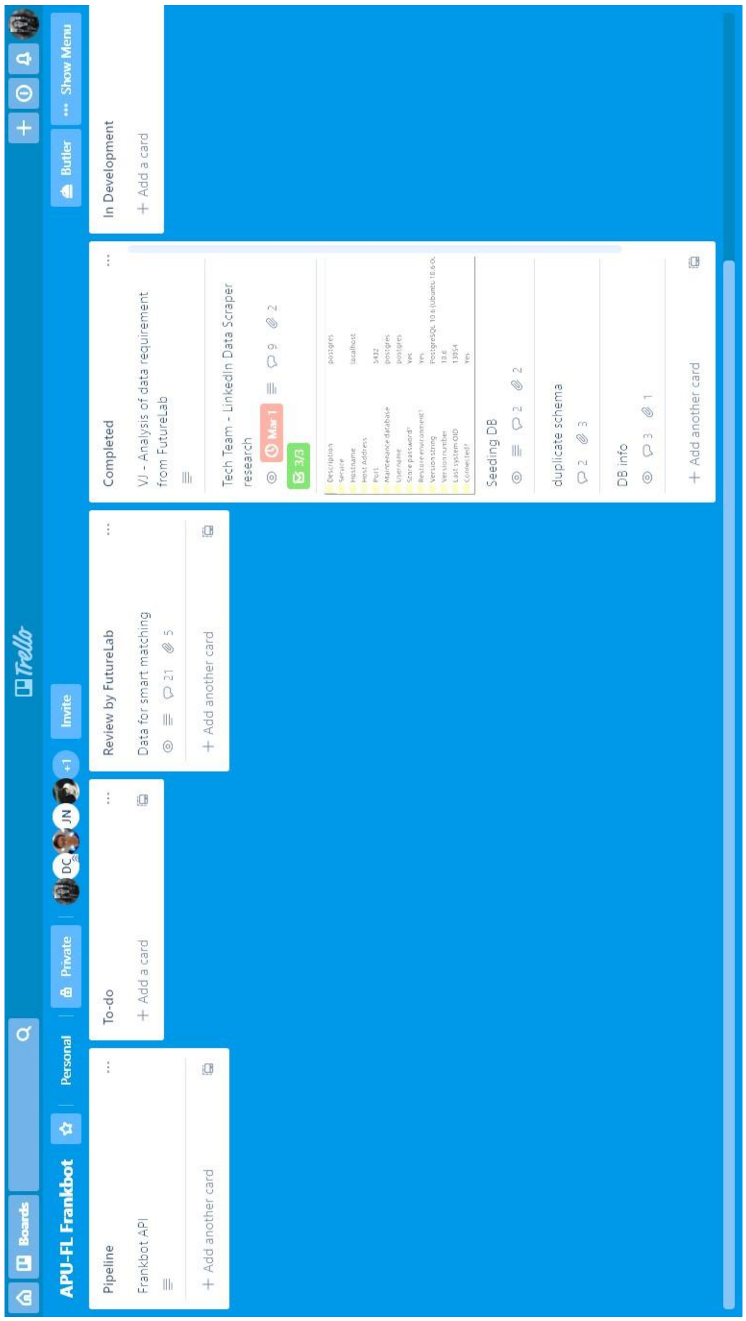

Figure 15 Project management dashboard as on 28-05-2019

SECTION 4

MODELLING AND INTERPRETATION

4.1 SOFTWARE & TOOL USAGE SPECIFICATIONS

The integrated development environment used in the recommender model development is Anaconda Jupyter Notebook, and the programming language used is python3. The data is fetched from the *.dump* file and is loaded to the Postgres server to view the .dump file. Once the required data tables, such as booking details of the mentees, are identified, it is then downloaded for data understanding which is the next step of the CRISP-DM methodology. The data is explored and preprocessed in the SAS Studio and Tableau. The data story is built in Tableau, enabling a better understanding of the data. The CSV data file from SAS studio is then used in building the recommender model. The model is then converted into a deployable file where it can be integrated based on the code development practice of FutureLab.

4.2 DATA EXPLORATION

The data exploration is carried out as a part of the data understanding step in the methodology. The initial data exploration consists of structural and statistical analysis of the data in the SAS Studio.After the data exploration, the data is prepared for modelling. This data file is uploaded to Tableau to create the dashboards. A data story is created to provide a high-level understanding of the data.

4.2.1 INITIAL DATA EXPLORATION

The SAS code used in this initial data exploration is attached to Appendix A. The steps followed, and output of initial data exploration are as follows:

1. Import the data into SAS Server – The data is uploaded into the SAS server from localhost

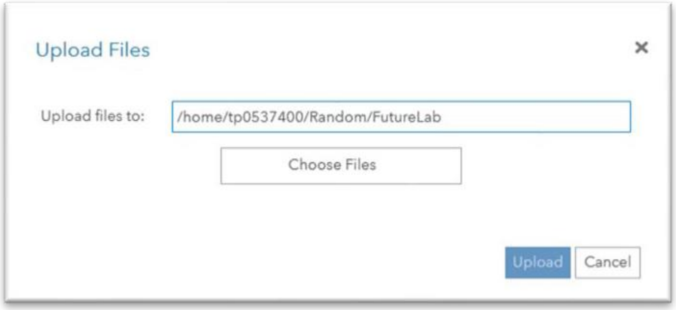

Figure 16 Import Data into SAS Server

2. Import the data into the SAS Studio working library - The data is loaded in the SAS studiolibrary.

```
NOTE: WORK.IMPORT data set was successfully created.
NOTE: The data set WORK.IMPORT has 2959 observations and 8 variables.
NOTE: PROCEDURE IMPORT used (Total process time):
      real time           0.15 seconds
```

Figure 17 Log of Data Import

3. Contents of the data: The summary of the contents and the dataset's variables is obtained.

The CONTENTS Procedure

Data Set Name	WORK.IMPORT	Observations	2959
Member Type	DATA	Variables	8
Engine	V9	Indexes	0
Created	12/20/2019 06:12:34	Observation Length	480
Last Modified	12/20/2019 06:12:34	Deleted Observations	0
Protection		Compressed	NO
Data Set Type		Sorted	NO
Label			
Data Representation	SOLARIS_X86_64, LINUX_X86_64, ALPHA_TRU64, LINUX_IA64		
Encoding	utf-8 Unicode (UTF-8)		

Figure 18 Dataset Contents

4. Statistics of the data – The means procedure is used to explore the dataset.

The MEANS Procedure

Variable	Minimum	Maximum	Mean	Median	Mode	Std Dev	N
mentor_id	1.0000000	8180.00	2590.36	2364.00	2985.00	1777.68	2959
mentor_count	1.0000000	124.0000000	30.0185874	14.0000000	5.0000000	34.5856082	2959
mentee_id	3.0000000	8225.00	3609.42	3439.00	1701.00	2153.03	2959
mentee_count	1.0000000	68.0000000	10.1922947	4.0000000	1.0000000	15.0122058	2959
mentor_score	0.5000000	4.3750000	3.0090825	3.0000000	3.0000000	0.3841938	2959
mentee_score	0.5000000	4.3750000	3.0092514	3.0000000	3.0000000	0.3840798	2959

The FREQ Procedure

mentor_industries	Frequency	Percent	Cumulative Frequency	Cumulative Percent
#customerexperience	2	0.07	2	0.07
#food science with nutrition	1	0.04	3	0.11
1. Course title- LAW LLB 2. Resume/CV Guidance 3. Malaysia/UK/US/Australia/___Universities Application	1	0.04	4	0.14
Account Management	2	0.07	6	0.21

Figure 19 Stats of Variables in Dataset

5. Univariate Analysis – The univariate analysis is used to obtain the statistical measures, extreme values and the following plots: It is seen that the observations have an average count of 3 in the mentor score.

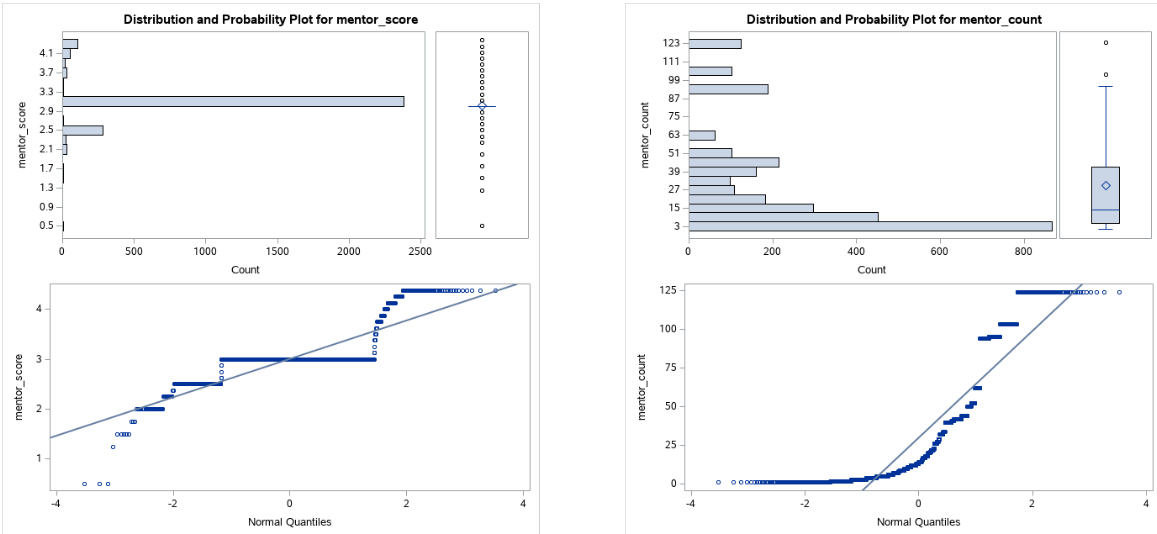

Figure 20 Univariate Analysis

6. Variable Correlation – The correlation table shows that both the variables have a positive correlation which means both the variables increase or decrease together.

Pearson Correlation Coefficients, N = 2959 Prob > |r| under H0: Rho=0		
	mentee_score	mentor_score
mentee_score	1.00000	0.99971 <.0001
mentor_score	0.99971 <.0001	1.00000

Figure 21 Ratings Scores Correlation

4.2.2 DATA PREPARATION

The previous section detects the missing values in the mentor and mentee rating score variables. As these variables are continuous, the mean imputation is carried out to treatthe missing values. It is not treated concerning outliers as extreme values are necessary for understanding the customer data. However, when it comes to recommender system output, theextreme values are neglected from sorting the mentor data.

4.2.3 DATA STORY

Data story is one of the most effective ways to communicate findings from the patterns of the data analysis. It has enabled an interactive experience in understanding the mentorand mentee data. The steps involved in creating a data story in Tableau started with creating individual charts, dashboards and finally, storyboards.

The individual charts are discussed here:

- *Visualisation #1 – Word cloud of mentor industries*. The bigger the word, the high is the frequency of the word. In this case, the word cloud shows the industries of mentors based on their most recent experience in the market. It is observed that Startups is the most repeated industry among the mentors, and Blockchain is the most demanded industry by mentees in the online mentoring website.

Mentor Industries

Mentee Interested Industries

Figure 22 Mentors' Industries (1)

- *Visualisation #2 – Treemap of mentor industries.* This visualisation shows the percentage of mentor's industries. A similar view is created for mentee interested industries also.

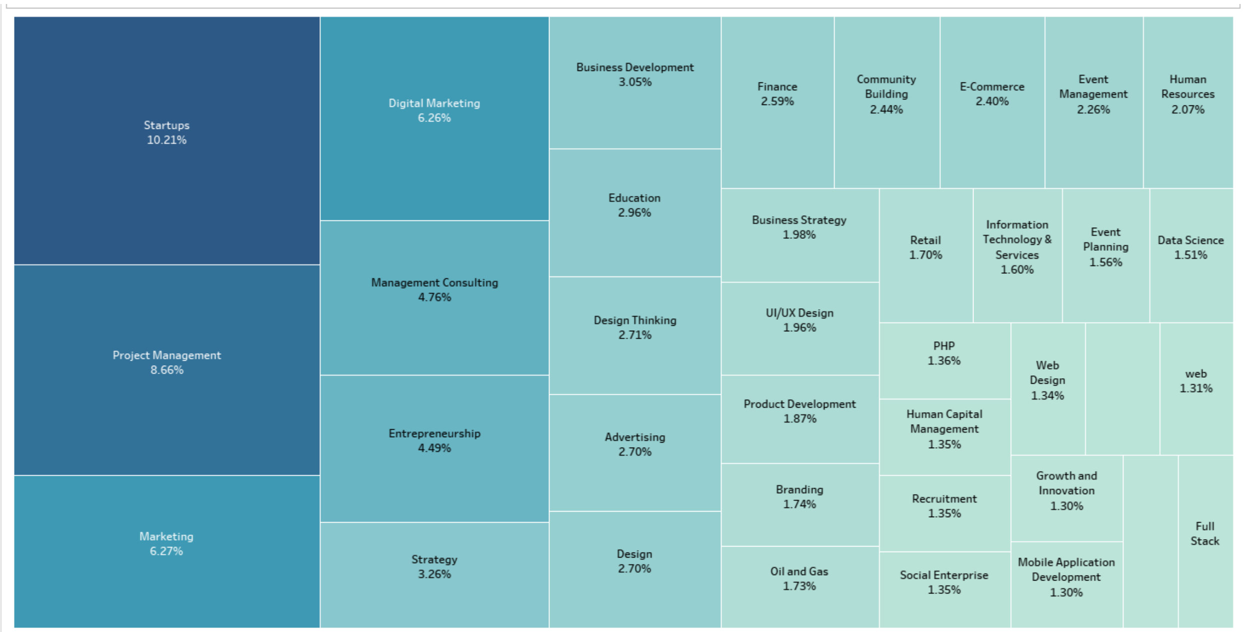

Figure 23Mentors' Industries (2)

- *Visualisation #3: Bar chart of Popular mentors and mentees.* This chart shows the popular mentor ID (1745) and popular mentee ID (5127). This can be used in the website testimonials.

Popular mentors ***Popular mentees***

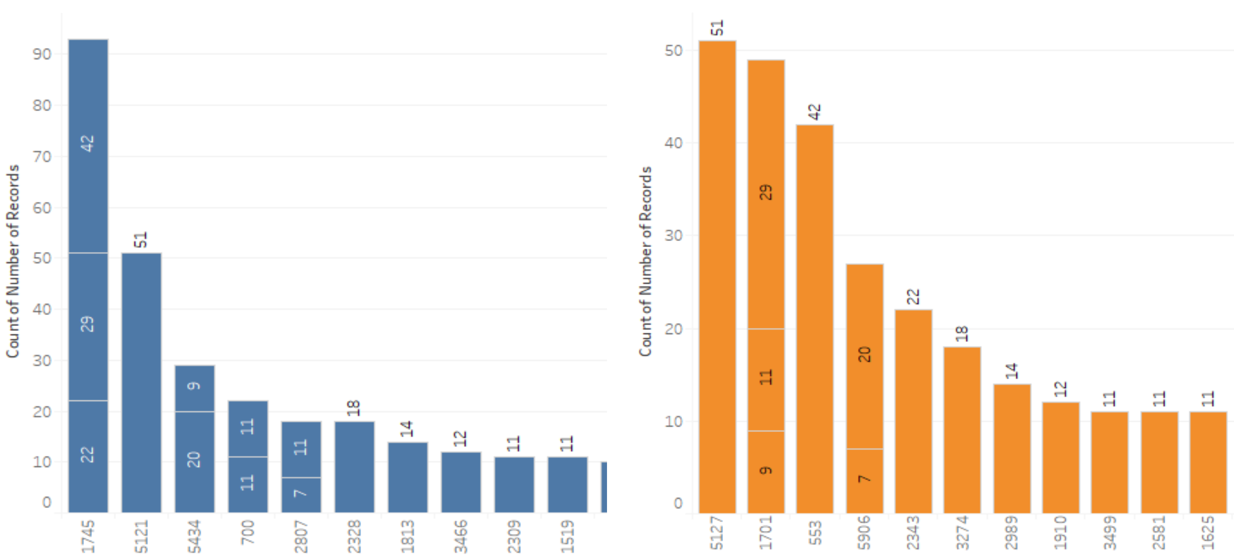

Figure 24 Popular mentor and mentee

- *Visualisation #4: Donut chart of Mentor and Mentee Rating Scores.* The mentors and mentees' ratings are shown in the below figure 25. It is observed that most of the customers are providing an average score of 3. Mentees have low scores (2), and mentorshave good scores in the upper quartile range, even 10.

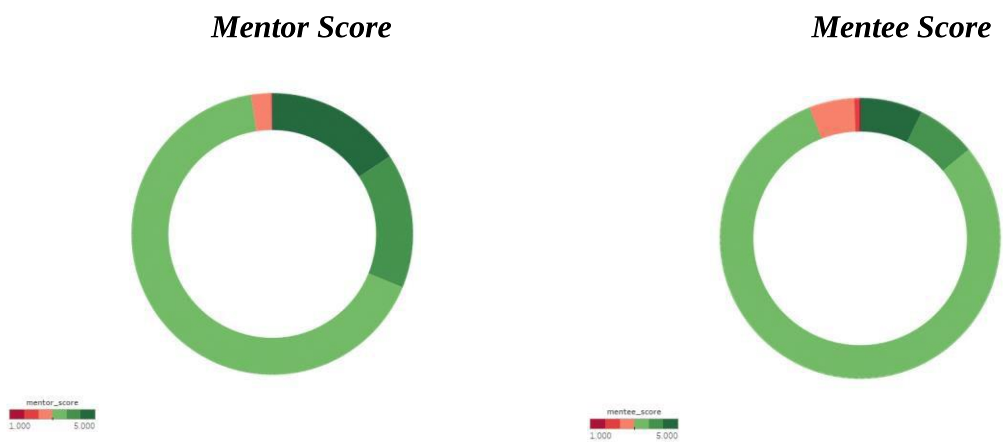

Figure 25 Mentor and Mentee Score

The visualisations are combined into dashboards. These dashboards are used to create the story cards in the storyboard. The snapshots of the storyboard are attached in Appendix A. This storyboard is used to present the data exploration. Connecting the real-time data source to this storyboard will act as the monitor system for customer behaviour in the application platform.

4.3 DEVELOPMENT OF ML RECOMMENDER SYSTEM MODEL

When the test website is loaded, it lands on the login page. When the customer is logged in, it loads the discover page. The mentors are suggested randomly, and there is a specific section called *'Go to Smart Matching'* as shown in figure 26. The recommender model is built to enhance the feature of smart matching in this website.

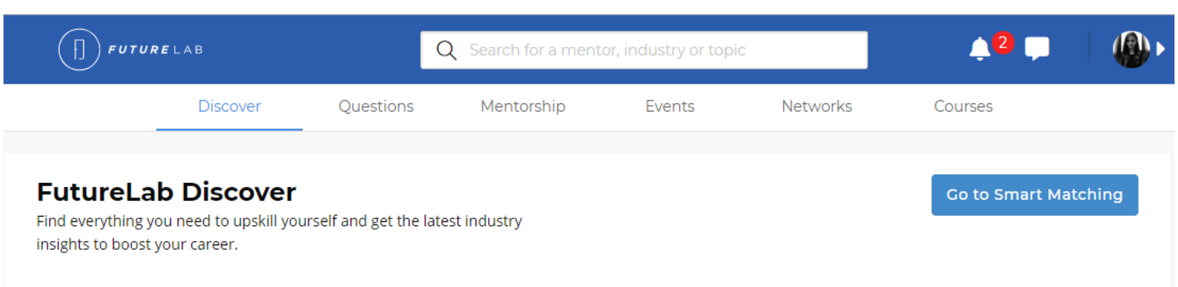

Figure 26 Smart Matching Function

4.3.1 STEPS INVOLVED IN BUILDING THE MODEL

Table 6 shows the steps followed in Jupyter notebook to build the recommendersystem model in python3. The libraries such as numpy, pandas, cosine_similarity (sklearn.metrics.pairwise), CountVectorizer (sklearn.feature_extraction) and joblib (sklearn.externals) are loaded before the first step.

Table 6 ML Model Building steps

Model Building steps	CRISP-DM Methodology	Library/ Function	Description
Step1: Read the file	Business understanding	`pandas.read_csv()`	Data is loaded after understanding the business
Step 2: Data Exploration	Data Understanding	`data.head()`	The data is explored in the Jupyter notebook.
Step 3: Data Transformation	Data Preparation	`cv=CountVectorizer()` `cv.transform()`	The data is vectorised.
Step 4: Cosine Similarity model	Model building	`cosine_similarity (inputUI_transform, datainDB_transform)`	This gives the array of index location (iloc) of the most similar data
Step 5: Model Testing	Evaluation	`Confusion_matrix()`	The values such as accuracy, Precision and recall can be calculated
Step 6: Packaging for deployment	Deployment	`joblib.dump()`	This creates the model file, where the model can be easily loaded.

The data is completely used for training as the new data will be used in testing the website. The model performance of the training data is evaluated in the following section.

4.3.2 MODEL PERFORMANCE

The trained model is manually tested for its performance. This is the evaluation step of the CRISP-DM methodology, where the accuracy, sensitivity, specificity and precision are calculated from the confusion matrix. The following table 7 shows the confusion matrix of the model where n is the number of test cases.

Table 7 Confusion Matrix

n=100	True Positive	True Negative	Total
Predicted Positive	53	21	74
Predicted Negative	12	14	26
Total	65	35	100

- Accuracy = (TP + TN) / (n) = 0.67 => 67%
- Sensitivity = TP / (TP + FN) = 0.81 => 81%
- Precision = TP / (TP + FP) = 0.71 => 71%
- Specificity = TN / (FP + TN) = 0.40 => 40%

The performance metrics of the model can be improved only by exposing the model to more and more data. As the model gets trained, the performance metrics will gradually increase. The model will now be tested on the website with testing data after the deployment.

4.3.3 OUTPUT SIMULATION & INTERPRETATION

The recommender system provides the mentor_id index location in an array. This array is sorted in the order of rating of the mentors. This can fetch the data from the location and populate the user interface. As the user enters the interested industry, the smart matchingis done by this model and can be fetched by the REST API.

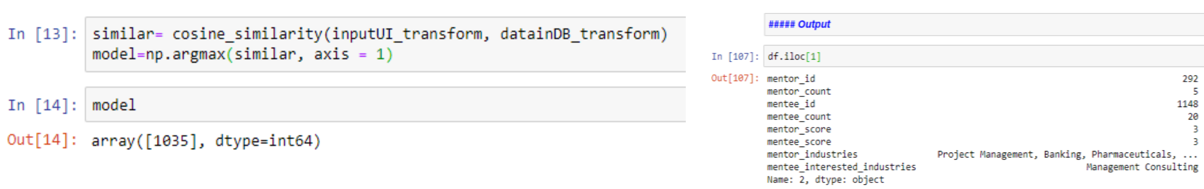

Figure 27 Output Simulation

4.4 NEW MODEL OF THE RATING SYSTEM

The existing model of rating the session by both the mentor and mentee consists of only the overall satisfaction and comments, which is not used in the feedback. Therefore, the recommender model is built incorporating the ratings of the mentor and mentee

4.4.1 PROPOSED MODEL FOR RATING

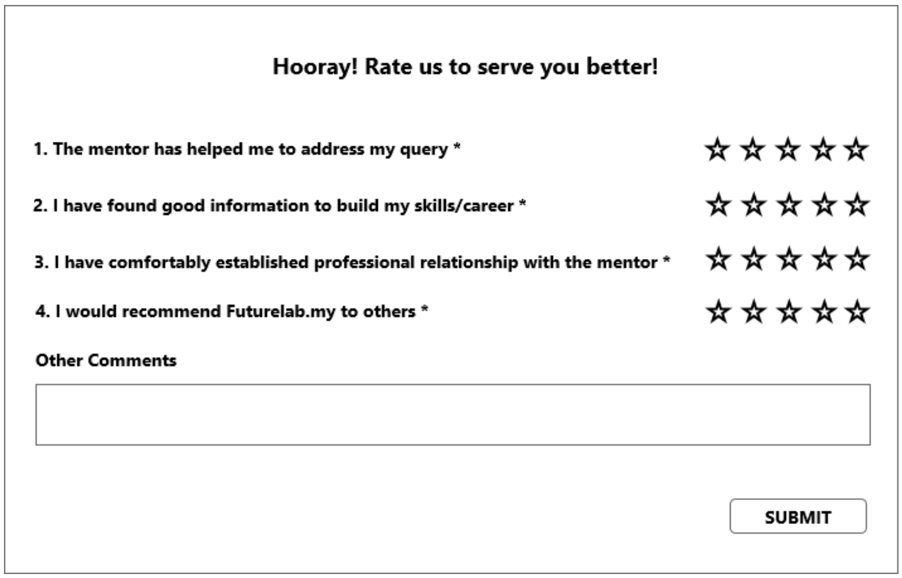

Figure 28 User Interface Design - Mentor's Rating by Mentees

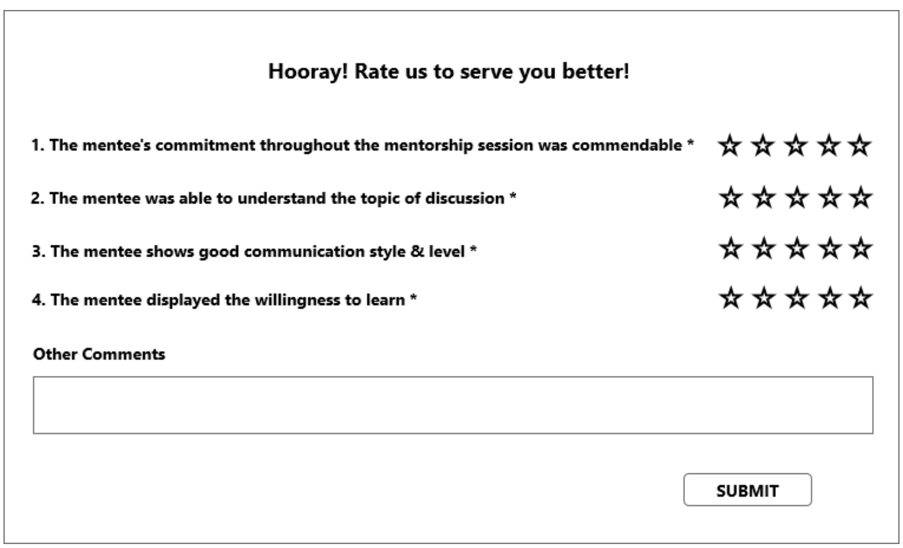

Figure 29 User Interface Design - Mentee's Rating by Mentors

The proposed model is different from the existing model, where the new model has four questions that reflects the commitment, communication, understanding and other metrics of both mentors and mentees. Other comments are also collected in this new model. The rating is done after the mentoring session is completed.

4.4.2 WORKING OF THE RATING MODEL

The model's working consists of after the session; mentee enters the rating and comments forthe mentors and mentors rate the mentees. Unlike the existing model, the new model has four questions. The average of the rating stars is calculated. This rating is then averaged into the mentor and mentee's overall score.

For example, assume that the mentor_id *17X8* has an overall score of 4. After today's session, the mentee has given the rating values of 3,4,4,3.5 for the four questions. Now, the average oftoday's session is (3+4+4+3.5)/4=3.6, and therefore overall score becomes (4+3.6)/2=3.7

4.4.3 RATING DATA COLLECTION AND INTEGRATION

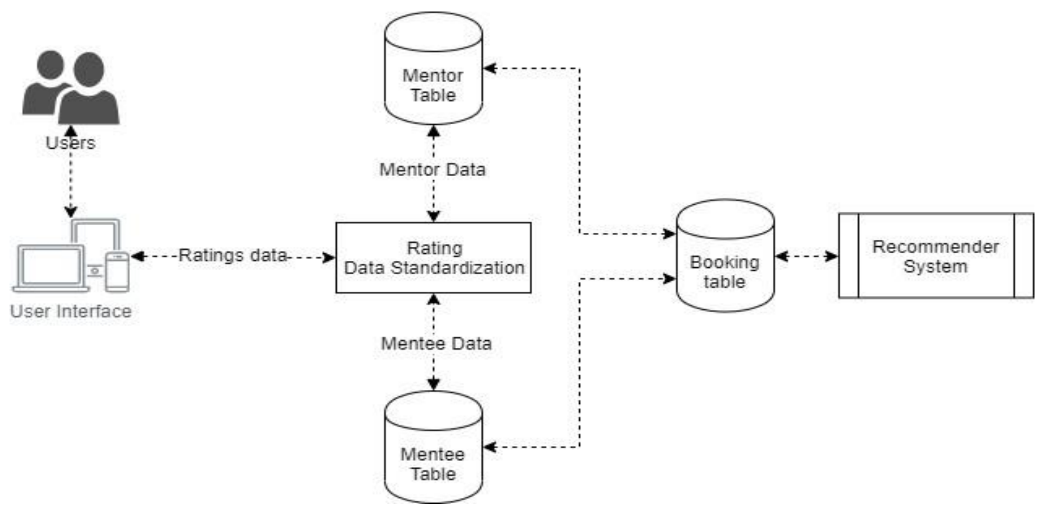

Figure 30 Ratings Data Collection and Integration

As the user enters the ratings, the data is standardised to calculate the overall rating and is then stored in the respective tables of mentor and mentees. When the recommender system data is fetched, these values are also reflected in the booking table. The comments are also saved in the respective tables of mentor and mentee. This overall score of mentors can be used to find the popular mentor in the cohort and populate the results on the discover page of the website.

4.5 MODEL DEPLOYMENT STEPS

The final step of the CRISP-DM methodology is deploying the model in the last stage of the data pipeline. The deployment of the model is done by loading the model accessible by the REST (REpresentational State Transfer) API (Application Programming Interface). It is highly recommended to maintain the managed deployment service pipeline (which consists of ingestion, transformation, model and deployment stages), enabling the continuous integration and deployment (CI/CD) services without any interruption.

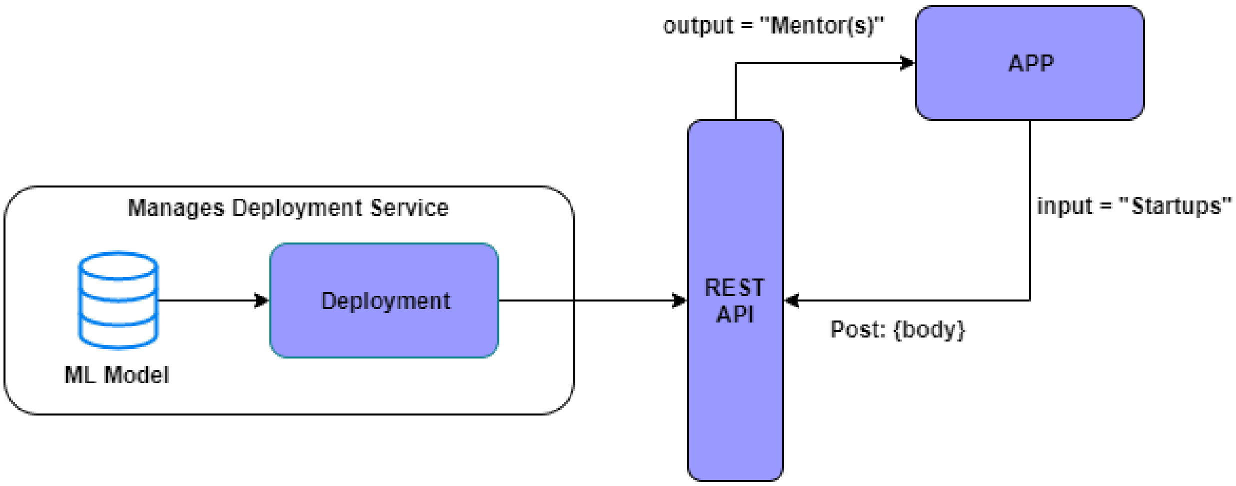

Figure 31 Deployment Steps

It is also recommended to connect the performance metrics and often update the data source into the data visualisation dashboards to enable the service monitoring feature. The cut-off values of the metrics can be set in these dashboards where the alarm is triggered when the value goes above/below the set value as per the requirement. This is an extensive step of deployment that covers the model maintenance.

It is important to consider how often the model needs maintenance or tuning. This completely depends on the model performance and feedback. The model needs to be tracked consistently from the log and monitor data. This can be automated or scheduled to maintain the optimum performance of the model. The ideal maintenance phase is carried out basedon the changes in the training data. Often, the maintenance is made to track records from the application and model activity logs.

SECTION 5
DISCUSSION AND CONCLUSION

5.1 DISCUSSION

The three main outcomes of this research are the user online behaviour model, recommender system and the rating model. The first outcome is one objective of this research, where the user behaviour is understood from the theoretical model. This model was developed keeping the online users in focus, where the information source is completely online. The behavioural model has a new level that emphasises the user's information-sharing behaviour. This level will guide the researchers on online customer behaviour and related projects.

Building the recommender system is another objective of this research, which was developed based on the booking data of the mentoring service. But before building this model, the important objective of identifying the parameters of the enterprise data will help the smart matching feature be identified. These parameters are then cleaned, preprocessed and prepared for the machinelearning model. Choosing this machine learning model was the outcome of understanding the business and its problem. The model is chosen considering the test website's upcoming features. It is then built and made ready for deployment in the pipeline.

Building the rating model is the final objective of this research. The complete rating system of the mentor and mentee is revamped into a new user interface and user experience. This model consists of questions that directly and indirectly reflect the user's characteristics based on their behavioural model. To rate the mentor, a mentee can choose the number of stars and add thecomments if needed. The same step is followed to add ratings and comments to the mentee.

5.1.1 RESULTS

The result status of each of the objectives and its outcome is checked in the following table 8:

Table 8 Results of the research

Objective No.	Objective	Outcome	Result Status
1	User online behavioural model	Theoretical Model	✓ Completed
2	Parameters of the model	Dataset	✓ Completed
3	Recommender model	Practical Model	✓ Completed
4	Ratings model	Practical Model	✓ Completed

5.1.2 OVERALL FINDING & ANALYSIS

An understanding of the business is made by the case study of the mentoring website, as shown in figure 32. This website is built to connect the potential talents and bridge theknowledge gap between mentors and mentees. As a web service, it opens up the volunteering platform for the mentors and paid subscription for the mentees. Our feature of the recommender system will fall in the section of Discover and Mentees page. The rating model will be incorporated at the mentoring session for both the mentor and mentee.

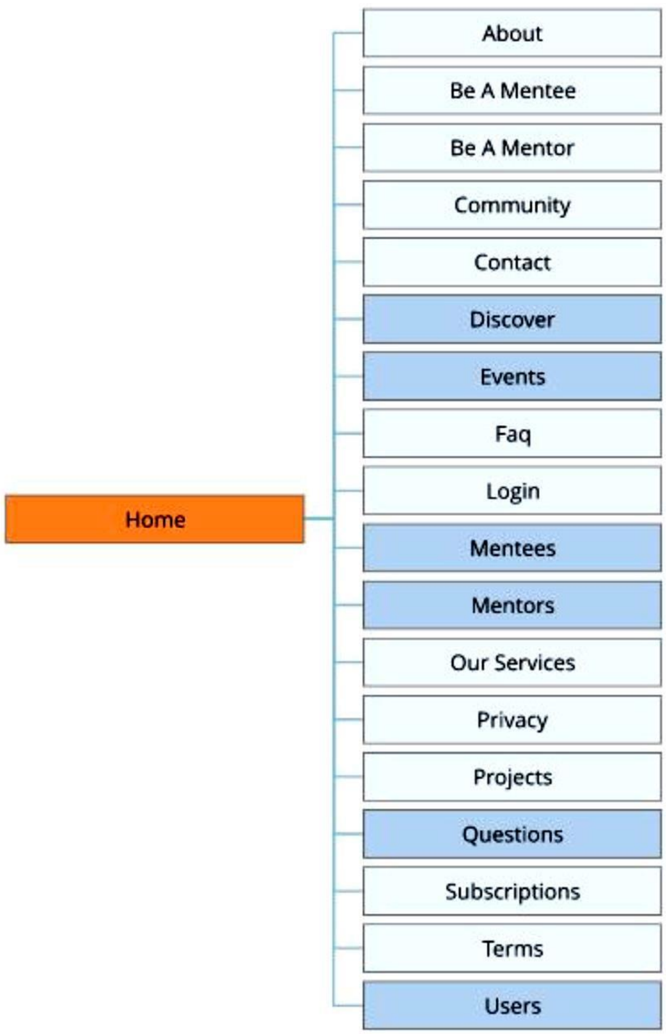

Figure 32 Case study of the online mentoring website

The mentoring session's data and user experience analysis was made from the interview data.The users have created a profile in the application, paid the subscription and took up the mentoring session on the website. A low participation rate and the lack of interest from mentors were observed in the analysis and the users' feedback. The common feedback was to revise the subscription, provide more details about mentors and suggest appropriate mentors. Our research focuses on solving the problem of suggesting the appropriate mentors.

5.2 MODEL IMPLICATIONS

The implications of this research work mainly outline the existing problem in the online mentoring website, such as understanding the user online behaviour, suggesting the mentors, maintaining the users' feedback, and providing an optimum solution to each of these problems. The theoreticaland practical implications are discussed in the following sections.

5.2.1 THEORETICAL IMPLICATION

The proposed theoretical model is extensively designed for users online behaviour, which is not covered in the existing behavioural models. The proposed model has a new level called information sharing behaviour. This model has helped understand the nuancesof the users' online behaviour. This can be followed in identifying the new feature for business development based on the users' behaviour and preferences.

5.2.2 PRACTICAL IMPLICATION

The trained machine learning model was developed mainly to recommend suitablementor profiles to the customers. The model has an accuracy of 67%, which can be improved by constant training. The recommendation model can be further optimised to improve the profile match by adding new features based on user behaviour. The machine learning models must be integrated and deployed in a continuous pipeline in the production environment. This will constantly improve the accuracy of the model.

Implementing the new rating model will improve the user experience and enhance the overall process of online mentorship. The rating model can identify the popular mentor to be featured on the front page of the application. It can also identify the less active menteeand provide target marketing such as discounts on the subscription fee.

5.3 RECOMMENDATIONS

The recommendations in a user interface, user experience and IT infrastructure are suggested based on the research analysis and feedback from the interviews of the user and the development team.

5.3.1 USER INTERFACE (UI)AND EXPERIENCE (UX) ENHANCEMENT

The following features are suggested for the improvement of UI/UX:

New Features	UI/UX	Description
Mentors before smart matching	UI/UX	This will help the users to choose the mentors based on popularity
Password reset	UX	The glitch in the Password reset by email needs to be fixed
Customer Support	UX	Extended customer support is required for moderation
Mentor Reviews	UI/UX	Reviews in the mentor profile will aid the user to choose the mentors
Choice of the payment	UX	This will improve the users booking experience
Mentors Profile	UI/UX	Reviewing the mentors' profiles needs to be done and also other professional profiles of the mentors must be made mandatory
Schedule	UX	Suggest alternative schedules tominimise the waiting time of the customer

5.3.2 IT INFRASTRUCTURE ENHANCEMENT

The recommendations in the information technology infrastructure are suggested to improve the model deployment. The migration of database and production platforms can be moved to the cloud-based architecture instead of server-based architecture. Cloud-based solutions provide scalability, reliability, easy integration through version control and automate the deployment process.

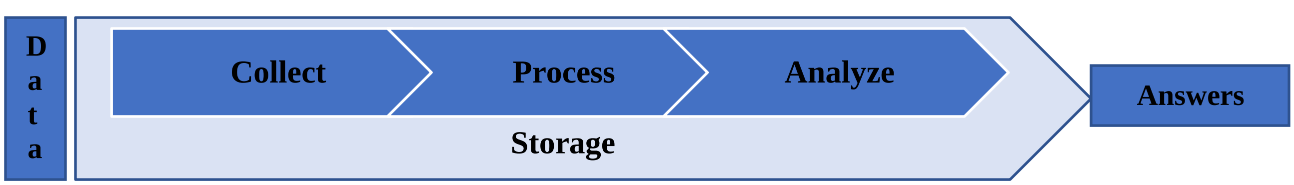

Figure 33 IT Infrastructure

5.4 CONCLUSION

The customers of the online mentoring platform, typically students, entrepreneurs, and working professionals, search for suitable mentors based on their profiles. This research was carried out to enhance the search feature to be the smart matching feature, which will provide themost suitable profile match to the customers. To understand the customer behaviour, their online search behaviour needs to be analysed. But the existing models are not emphasising online behaviour. After researching existing models, the user online search behaviour is defined in theproposed model. The theoretical implications of the behavioural model cover all the possible factors and levels of online user behaviour.

Understanding the business and data helped in choosing the parameters of the model dataset. As the dataset is formed from the overall enterprise data, the modelling and designing of an optimum recommender system for the chosen online mentoring service are successfully built and trained. Thus, all the objectives of the research are met. This machine learning model is then converted into the model file, which can build the application programming interface (API). ThisAPI can be used in testing the model in production. After the deployment in the production environment, the users can use the feature of smart matching.

5.5 FUTURE WORK

The integration and deployment of the machine learning models have to be made continuous so that the data is completely utilised in providing better and uninterrupted service. Future work can be carried out in tuning the model based on the test data, which is important as it improves the model performance metrics. It can be enhanced by building an ensemble model of various featureswhich will be integrated into the data pipeline of the service.

REFERENCES

Awasthy, D., Banerjee, A., & Banerjee, B. (2016). Understanding the role of prior product knowledge to information search: An application of process theory to the Indian market. *Asia Pacific Journal of Marketing and Logistics*, *24*(2), 257–287. https://doi.org/10.1108/13555851211218057

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

Brett, J. (2018). Career Success Pyramid. In *Evolving Digital Leadership* (pp. 53–61). Berkeley, CA: Apress. https://doi.org/10.1007/978-1-4842-3606-2_4

Butner, K., & Ho, G. (2019). How the human-machine interchange will transform business operations. *Strategy and Leadership*, *47*(2), 25–33. https://doi.org/10.1108/SL-01-2019-0003

Caeser, V. (2018). Content Based Recommender System. Retrieved from https://www.kaggle.com/varian97/content-based-recommender-system#Content-Based-Recommender-System

Chatterji, A., Delecourt, S. M., Sharique Has, & Koning, R. (2018). *When Does Advice Impact Startup Performance?* https://doi.org/10.1002/smj.2987

Clevers, H. (2018). Mentoring the Next Generation. *Cell Stem Cell*, *23*(4), 468–470. https://doi.org/10.1016/j.stem.2018.09.014

Dixit, S., Badgaiyan, A., & Khare, A. (2017). An integrated model for predicting consumer's intention to write online reviews. *Journal of Retailing and Consumer Services*, (October), 0–1. https://doi.org/10.1016/j.jretconser.2017.10.001

Dörnyei, K. R., & Gyulavári, T. (2016). Why do not you read the label? - an integrated framework of consumer label information search. *International Journal of Consumer Studies*, *40*(1), 92–100. https://doi.org/10.1111/ijcs.12218

Egri, G., & Bayrak, C. (2014). The role of search engine optimisation on keeping the user on the site. *Procedia Computer Science*, *36*(C), 335–342. https://doi.org/10.1016/j.procs.2014.09.102

Ellis, D., & Haugan, M. (1997). Modelling the information seeking patterns of engineers and research scientists in an industrial environment. *Journal of Documentation*, *53*(4), 384–403. https://doi.org/10.1108/EUM0000000007204

FutureLab. (2019). Company website. Retrieved August 20, 2019, from https://futurelab.my

Gottlieb, M., Fant, A., King, A., Messman, A., Robinson, D., Carmelli, G., & Sherbino, J. (2017). One Click Away: Digital Mentorship in the Modern Era. *Cureus*. https://doi.org/10.7759/cureus.1838

Kaur, G., & Kaur, S. (2008). *MALAYSIAN GRADUATES' EMPLOYABILITY SKILLS*. *UNITAR E-JOURNAL* (Vol. 4). Retrieved from https://s3.amazonaws.com/academia.edu.documents/32462701/GurvinderMalaysianGraduat e_1.pdf?AWSAccessKeyId=AKIAIWOWYYGZ2Y53UL3A&Expires=1552857628&Signature=E%2FBLuXHAJCK7y%2F6xk7rT6Kfj4QU%3D&response-content-disposition=inline%3B filename%3DMALAYSIAN_GRADU

Kumar, S., & Johnson, M. (2017). Online mentoring of dissertations: the role of structure and support. *Studies in Higher Education*, *0*(0), 1–13. https://doi.org/10.1080/03075079.2017.1337736

Leckie, G. J., Pettigrew, K. E., & Sylvain, C. (1996). Modeling the Information Seeking of Professionals: A General Model Derived from Research on Engineers, Health Care Professionals, and Lawyers, *66*(2), 161–193. https://doi.org/10.1086/602864

Li, A., & Guan, Z. (2013). Your Search Behavior and Your Personality (Vol. 7719). https://doi.org/10.1007/978-3-642-37015-1

Linden, G., Smith, B., & York, J. (2018). [CF] Amazon Recommendations Item-to-Item Collaborative Filtering (Amazon 2003). *Mississippi Legislature*, (February), 2. https://doi.org/10.1038/sj.onc.1203797

MacKinnon, R. (2012). The Netizen. *Development (Basingstoke)*, *55*(2), 201–204. https://doi.org/10.1057/dev.2012.5

Martin, F., & Bolliger, D. U. (2018). Engagement Matters: Student Perceptions on the Importance of Engagement Strategies in the Online Learning Environment. *Online Learning*, *22*(1). https://doi.org/10.24059/olj.v22i1.1092

Pillon, S., & Osmun, W. E. (2013). Mentoring in a digital age. *Canadian Family Physician*, *59*(4).

Prendergast, G., Paliwal, A., & Chan, K. K. F. (2018). Trust in online recommendations: an evolutionary psychology perspective. *International Journal of Advertising*, *37*(2), 199–216. https://doi.org/10.1080/02650487.2016.1239879

Reddy, B. S., Krishnamurthy, M., & Asundi, A. Y. (2018). Information use, user, user needs and seeking behaviour: A review. *DESIDOC Journal of Library and Information Technology*, *38*(2), 82–87. https://doi.org/10.14429/djlit.38.2.12098

Robson, A., & Robinson, L. (2013a). *Building on models of information behaviour: Linking information seeking and communication*. *Journal of Documentation* (Vol. 69). https://doi.org/10.1108/00220411311300039

Robson, A., & Robinson, L. (2013b). Building on models of information behaviour: Linking information seeking and communication. *Journal of Documentation*, *69*(2), 169–193. https://doi.org/10.1108/00220411311300039

Solihin, N. (2013). Search Engine Optimisation : A Survey of Current Best Practices Search Engine Optimization : A Survey of Current Best Practices By. https://doi.org/10.1109/COMPSAC.2017.119

Stack, S. (2015). Learning Outcomes Online vs. Traditional courses.pdf, *9*(1).

Tury, S., Robinson, L., & Bawden, D. (2015). The Information Seeking Behaviour of Distance

Learners: A Case Study of the University of London International Programmes. *Journal of Academic Librarianship*, *41*(3), 312–321. https://doi.org/10.1016/j.acalib.2015.03.008

Utkarsh, Sangwan, S., & Agarwal, P. (2018, January). Effect of consumer self-confidence on information search and dissemination: Mediating role of subjective knowledge. *International Journal of Consumer Studies*, pp. 46–57. https://doi.org/10.1111/ijcs.12482

Valaei, N., & Rezaei, S. (2016). Job satisfaction and organisational commitment: An empirical investigation among ICT-SMEs. *Management Research Review*, *39*(12), 1663–1694. https://doi.org/10.1108/MRR-09-2015-0216

Wang, M. (2007). Designing online courses that effectively engage learners from diverse cultural backgrounds. *British Journal of Educational Technology*, *38*(2), 294–311. https://doi.org/10.1111/j.1467-8535.2006.00626.x

Wang, Y.-R., & Gibson, G. E. (2017). Pre-Project Planning and its Practice in Industry. *Proceedings of the 23rd International Symposium on Automation and Robotics in Construction*, *21*, 89–95. https://doi.org/10.22260/isarc2006/0161

Wilson, T. D. (1999). Models in information behaviour research. *Journal of Documentation*, *55*(3), 249–270. https://doi.org/10.1108/eb026677

Wirth, R. (2000). CRISP-DM : Towards a Standard Process Model for Data Mining. *Proceedings of the Fourth International Conference on the Practical Application of Knowledge Discovery and Data Mining*, (24959), 29–39. https://doi.org/10.1.1.198.5133