Behavioral Targeting:
Early Student Profile Identification for Efficient Student Recruitment

Master Thesis

Student: Yaqi Yin
Student number: s1469924

Faculty: Behavioral, Management and Social Sciences
Programme: Business Administration MSc
Specialization: Strategic Marketing and Business Information

1st Supervisor: Dr. E. Constantinides
2nd Supervisor: Dr. S.A. de Vries

Date: 1st August, 2018

Faculty of Behavioural, Management, and Social Science
University of Twente
P.O. Box 217
7500 AE Enschede
The Netherlands

UNIVERSITEIT TWENTE.

Abstract

With the rapid growth of online activities more marketers have switched their attentions to target the potential customer through the internet. As one of the growing popular online targeting methods, behavior targeting has been increasingly used to identify the potential customers based on their historical online activities. However, higher education industry has not utilized behavioral targeting in student recruitment. In addition, current studies mainly emphasize the effectiveness of behavioral targeting and focus on the design of behavioral targeting techniques instead of applying behavioral targeting to solve specific problems.

In this research, behavioral targeting has been extended to the higher education sector. The outcomes of this study explore online behavioral profiles of the students who searched master related online information and submitted online university applications from 2016 to 2017. The source of the data is from the dataset of the department marketing and communication in university of Twente. In order to analyze the online behaviors, a behavioral targeting analysis approach with the use of machine learning has been developed. It consists of a two-step clustering model and user characteristics description to assist in targeting the prospective students. This approach could be generalized to use in other relevant future research.

Key words: user profiling, behavioral targeting, clustering analysis

Table of Contents

1. Introduction

Nowadays, with the rapid grown of World Wide Web (WWW), the resource of online data has been utilized by the marketers to target the potential customers (Amoretti, Belli, & Zanichelli, 2017). There are three types of online targeting methods which are commonly used to identify the potential online customers (Pandey, et al., 2011). The first type of targeting method is property targeting. It delivers the ads on certain websites where the targeting customers will visit, such as placing basketball ads on the websites about the sports. However, the pitfall of this method is that it may lose the users who use other websites to search for the related information. As for the second targeting method, user segment targeting is applied to segment the customers based on demographic, geographic and psychographic attributes (Pandey, et al., 2011; Hamka, Bouwman, de Reuver, & Kroesen, 2014). However, the segmentations are pre-defined. It means that the segmentations are formed without testing and confirming by the real market. Moreover, the user segment targeting is only adequate to targeting broad groups based on the above three attributes. Thus, this popular method has a risk that the pre-defined segments may not be correct and accurate enough to capture the potential customers. The third type of targeting method is introduced called behavioral targeting. It refers to an approach where the potential customers are targeted based on their historical online behavioral data (Lu, Zhao, & Xue, 2016). In behavioral targeting, users are identified according to their online behaviors in which similar online behaviors lead to similar interests of users (Gong, Guo, Zhang, He, & Zhou, 2013). Yan et al. (2009) found once behavioral targeting was applied in user segmentation according to past browsing and searching behaviors, the effectiveness of the advertisement could increase by as much as 670%. Comparing with the other two targeting methods, behavioral targeting can provide a comprehensive and accurate targeting (Fan, Chow, & Xu, 2016). Hence, in order to have an accurate customer targeting, behavioral targeting is introduced in this research.

Before targeting the potential customers, online user profile is necessary to explore as a premise. It means that the marketers need to use the online user profile to find potential customers. There are two components in the online use profile which are behavioral profile and users' static characteristics profile (Dam & de Velden, 2015). Behavioral profile represents the dynamic online behavioral patterns of the potential customers. Users' static characteristics profile records the demographic, geographic and psychologic characteristics of the users in the behavioral profile. Hence, both profiles are required in this research. In addition, the hierarchy of effects model is applied to clarify the customer conversion funnel from the awareness to final purchase (Ghirvu, 2013). This model is adapted to this research to explain the different clusters of online behavioral patterns.

1.1. Research Gap

The above studies about behavioral targeting have stated its advanced advantages comparing to the other targeting methods. However, the behavioral targeting has not be applied in the higher education sector to assisting in student profiling identification. Besides that, the analysis approach to explore the student profiling based on their online behaviors has not been researched yet as well. Hence, there is a research gap in investigating behavioral targeting in higher education industry to guide the universities to increase the submitted study application and therefore, to increase the student recruitment eventually.

1.2. Objective and Research Question

The objective of this research is to explore online behavioral profiles of the students who searched master related online information and submitted online university application from 2016 to 2017 and to develop a behavioral targeting analysis approach with the use of machine learning. The research question is formed as follows:

'*What are the different online behavioral profiles of students based on master related online search information and submitted online university applications?'*

In order to answer the research question, the following sub-questions have to be solved:

1. What is an online behavioral profiling?

2. How does the profiles be identified?

3. What are the students' static characteristics in the behavioral profile?

1.3. Methodology

The targeted group in this research are the students who have searched information about master studies and submitted their study application through the website of university of Twente from July 2016 to August 2017. Two types of data extracted from the targeted students are used in this study which are web data and CRM data. The data in this research are accordance with the GDPR, thus it is ethical to apply these two types of data in this research.

A quantitative approach is conducted in this study. As the raw data are extracted as categorical data, the multiple correspondence analysis (MCA) has been applied to transfer the categorical data into continuous data. Based on the objective and the research question of this study, a behavioral targeting analysis approach is developed. It consists of two parts. The first part is cluster analysis with a two-step clustering model used as the multivariate technique. The cluster analysis is conducted in R software. Another part of the approach is user characteristics description with the use of IBM Watson Analytics.

1.4. Implication

When it comes to the implication, from the perspective of theoretical implication, this research has developed an analysis approach about exploring the behavioral profiles. Moreover, the research extends the scope of the Lavidge and Steiner hierarchy of effects model to assist the university in students' recruitments. As for the practical contributions, first, the behavior profiles could assist the university placing the online advertising on the webpages where the students' browser, visit, click and download frequently. In addition, combining with student characteristic profile, the university could have a precise targeting of prospective students.

1.6. Outline of the study

In the first section, an introduction of this study is presented. In the second section, the theoretical background about use profiling, behavioral targeting, machine learning, and hierarchy of effects model are explained. In the methodology section, data description, measurement and data analysis are studied. In the next section, there is an indication about the result of this research. Next, conclusion and discussion are explained. This study will end up with the sector of limitation and future research.

2. Literature Review

In this section, core literatures used in this research are clarified. The core literatures explain various aspects such as user profiling, behavioral targeting, machine learning, and the hierarchy of effects model. Each sub-section performs a particular aspect starting from the definition to contents and ending with how it adapts to this research.

2.1. Online user profiling

2.1.1. What is online user profiling?

User Profiling refers to a process of identifying the data about a user's interest domain (Kanoje, Girase, & Mukhopadhyay , 2014). Different from the definition of traditional user profiling, the concept of online user profiling emphases more on users' online behaviors. Online user profiling is defined as a practice of tracking information about the consumers by monitoring their online movements. It contains all actions associated with obtaining, studying and applying data related to user behavior within a network (Al-Qurishi, et al., 2018). As the contents of online user profiling, online user profile is a summary of a consumer's interests and preferences revealed through the consumer's online activity (Trusov & Ma , 2016).

2.1.2. The dimension of online user profiling

The online user profiling approach is clarified in three dimensions which are:

- Online user profiling contents: static profiling, dynamic profiling
  (Kanoje, Girase, & Mukhopadhyay , 2014; Dam & de Velden, 2015; Krishnan & Kamath , 2017; Amoretti, Belli, & Zanichelli, 2017)
- Online use profiling acquisition approaches: explicit profiling, implicit profiling
  (Cufoglu , 2014; Kanoje, Girase, & Mukhopadhyay , 2014; Hawalah & Faslia, 2015; Amoretti, Belli, & Zanichelli, 2017; Arzubov, Shakhovska, & Lipinski, 2017)
- Online user profiling filtering methods: Collaborative
  (Poo, Chng, & Goh, 2003; Cufoglu , 2014; Hawalah & Faslia, 2015; Fan, Chow, & Xu, 2016)

<u>Online user profiling contents</u>

Static profiling refers to a process of analysis about the users' stable characteristics that rarely never change (Krishnan & Kamath , 2017). In the static profiling, the users' basic demographic and geographic information is presented such as name, age, country, email address and so on. Another type of profiling is called *dynamic profiling*. It is defined as a process of analysis about the users' online activities to target the interests of users. The dynamic profiling consists the users' online behaviors and captures the user' interests which is also called behavioral profiling (Amoretti, Belli, & Zanichelli, 2017). This type of profiling is frequently updated on the basis of online browser history. The reason behind is that new online activities always occur and the customer' interests changes frequently (Krishnan & Kamath , 2017). Hence, it contributes dynamic and temporal attributes to the online user profile.

Online use profiling acquisition approaches

There are two types of online user profiling acquisition approaches: explicit and implicit. According to Kanoje et al. (2014) *explicit profiling* represents acquiring the user's information in explicit ways by asking the visitors to fill in the forms and selection menus before or during visiting the webpages. In other words, users actively choose to create their own user profiles when filling in the questionnaires, surveys and (online) forms in explicit profiling (Cufoglu , 2014). In addition, combining with the dimension of user profiling contents, the static profiling is commonly acquired with the explicit acquisition approach (Dam & de Velden, 2015). The advantage is that the process of gathering the static information is efficient in the explicit acquisition approach. However, the pitfall of the profiling approach is that concerning about their own privacy or preferences, it could be possible that some users try to avoid revealing their information when filling in the questionnaires and forms. Thus, it influences the accuracy of the information in their online user profile. While *implicit profiling* refers to gathering the information by observing and monitoring the users' interactions with automatic systems (Kanoje et al., 2014). The implicit approach is used to acquire dynamic information by analyzing the behavioral patterns of the users with the technique of machine learning algorithms (Arzubov, Shakhovska, & Lipinski, 2017). However, the implicit acquisition approach has a withdraw that is the customer privacy concerns. Several previous studies (Debatin & Lovejoy , 2009; Goldfarb & Tucker , 2011; Caudill & Murphy, 2012) have worried about that the visitors do not have the awareness about the extents of online tracking when having web usage activities which threats the visitors' privacy. In order to solve privacy problems, consummate privacy regulations and policies have been developed. The regulations standardize the marketers' monitor and tracking behaviors to protect the customers' legal rights (Dam & de Velden, 2015).

Online use profiling filtering method

Yang et al. (2017) states that *collaborative-based method* is the idea of extracting information by learning from the users who have the similar behaviors and preferences. In addition, the extracted information is grouped into different clusters based on their similar behavioral patterns with the clustering technique (Algiriyage , Jayasena , & Dias, 2015; Mustafal, Ibrahim, Ahmed, & Abdullah, 2017). The procedure of the clustering techniques is explained in detail in section 2.3.

2.1.3. Adaption to this research

A hybrid online user profiling approach is conducted in this research. The dynamic, implicit profiling with collaborative filtering method is used to group the students into different clusters. It indicates that web data in this research are gathered by tracking and storing the user' online behaviors in the website of the university (dynamic and implicit). In addition, the students who have performed similar online behavioral patterns are grouped into one user behavioral profile (collaborative). After clustering the students, combining with the students' static profiling gathered from customer relationship (CRM) database, the online user profile for each cluster is developed.

2.2. Behavioral Targeting

Online user profiling is a cornerstone of behavioral targeting. In other words, the objective of online user profiling is to conduct behavioral targeting (Trusov & Ma , 2016).

2.2.1. What is behavioral targeting?

Behavioral targeting is defined as an approach that delivering the information to the targeting users based on their historical online behavioral data (Lu, Zhao, & Xue, 2016). The objective of behavioral targeting is to identify the users who are more interested in the offered online information. In addition, there are two features of behavioral targeting: tracking users' online behaviors, and using the collected behavioral data to target customers.

2.2.2. Types of online behavioral data

Online behaviors have web browsing data, search history data, purchases data, downloading data, click through responses to ads (ad clicks) data and communication contents. They can be concluded as browsing activities (web page visits) and search query (Pandey et.al, 2011; M ,Thottungal, & Nizar, 2016; Boerman, Kruikemeier, & Borgesius, 2017).

In order to track the browsing and searching activities, "cookies" have been commonly used to record and monitor the web visiting history and online actions of the users in details (Chen & Stallaert, 2014). Cookies are small text files that a server can send to a web visitor (Borgesius, 2015). In addition, Arzubov et al. (2017) states that a browser cookie is set by the website publisher who owns the domain to which the cookies belongs. This type of cookie is also called 'first-party cookie'. When the browser cookie is set by the third party instead of the website publisher, this type of cookie is called 'third-party cookie'. The users saves the cookies in the computer when accessing to a single website. If the user visits the website and has online interactions with the server again, the server will recognize this web user by the cookies stored on the user's computer. Thus, the online users' behaviors are gathered and tracked by the website publisher continuously.

2.2.3. Types of behavioral targeting

Srimani & Srinivas (2011) found that there are two types of behavioral targeting: ad network behavior targeting and on site behavioral targeting.

- Ad network behavior targeting means the advertisements have lots of online publishing platforms. The visitors who have online interactions on the ad could be regarded as having the similar interests. For instance, the advertiser places the advertisements on different online platforms, such as Facebook, YouTube, Twitter and Instagram. The visitors who have interactions with the advertisements on these platforms are treated as the potential customers with similar interests.
- On site behavioral targeting refers to targeting the potential customers based on the within-site browsing activities of users at a (commercial) website (Fan,

Chow, & Xu, 2016). The web visitors are segmented into different groups according to their online browsing behaviors on that certain website with clustering algorithm.

2.2.4. Behavioral Targeting Performance Measurement

The study from Pandey et al. (2013) defines conversion as the direct measurement of behavioral targeting performance. Besides using as evaluation criteria, reaching the estimated conversion is also the objective of the behavioral targeting (Pandey, et al., 2011). When having the behavior targeting, we could choose different criteria as conversion based on their own objectives (Bagherjeiran, Hatch, & Ratnaparkhi , 2010).

2.2.5. Adaption to this research

On site behavior targeting will be used for user profiles. It means that the research focuses on the online behaviors conducted on a single website (the official website of university of Twente). The students' online behavioral data on the website of university of Twente are tracked by the third-party cookie. Based on the objective of this research the conversion is defined as the number of study application submission.

After collecting and extracting online behavioral data, the online behavioral patterns need to be explored with the collaborative filtering method. As the primary mechanism, machine learning is applied in this research to explore the online behavioral patterns.

2.3. Machine learning

2.3.1. What is machine learning?

Machine learning is defined as the process of automatically discovering patterns and trends in data that go beyond simple analysis (Ghatak , 2017). It has become a primary mechanism for extracting the structured information and knowledge from raw unstructured big data. The structured information and knowledge are transformed as automatic predictions for various applications (Xing, Ho, Xie, & Wei, 2016). There are diverse applications of the machine learning including user profiling, behavioral targeting, data mining, recognition systems and recommendation systems (Qiu, Wu, Wu, & Feng, 2016).

Machine learning can be performed in supervised and unsupervised ways. *Supervised machine learning* requires training the data that has the input and desired output (Qiu et al.,2016). When having the training data and desired output, the supervised machine learning algorithms will learn and use the knowledge to classify the new data (Portugal, Alencar, & Cowan, 2018). In the recent literature reviews from the last 3 years, two subcategories of supervised machine learning are classification and regression (Gupta, Sharma, & Jindal, 2016; Singh, Thakur, & Sharma, 2016; Portugal et al., 2018). *Unsupervised machine learning* involves analyzing unlabeled input data. Unsupervised machine learning does not require labeled training and desired output (Qiu et al.,2016). Unsupervised machine learning is commonly applied to find out hidden patterns in the raw data and make predictions (Portugal et al., 2018). As one of the important types of unsupervised machine learning, clustering algorithm is applied to perform customer segmentation. Based on the literature review of Portugal et al. (2018), the collaborative filtering technique with clustering in online user profiling is the one that has been most used.

2.3.2. Clustering method

Clustering is a machine learning algorithm that splits multivariate data into different groups based on similar features of the data (Bang, Cho, & Kim, 2015). In this research we focus on two types of clustering methods: hierarchical clustering and k-means clustering. The literature review about user profiling with clustering algorithms in the past 4 years are shown in table 1.

Table 1. Literature Review about user profiling with clustering algorithms

Clustering algorithm	Author
Hierarchical clustering	Algiriyage, Jayasena, & Dias, 2015; Kumar & Ashraf, 2015
K-means	Bang, Cho, & Kim, 2015; Kumar & Ashraf, 2015; Qin, Guan, Wang, & Liu, 2015; Zahra, Ghazanfar, Khalid, & Azam, 2015; Xiu, Lan, Wu, & Lang, 2017
MCA K-means	Dam & de Velden, 2015; Mitsuhiro & Yadohisa, 2015

Hierarchical clustering

Hierarchical clustering algorithm uses either top-down approach (divisive clustering) or bottom-up approach (agglomerative clustering). Agglomerative clustering is more common used in user profiling to perform behavioral targeting (Balcan, Liang, & Gupta, 2014; Kumar & Ashraf, 2015). Agglomerative clustering merges the objects (or groups) that are close to each other in the same cluster. The termination is triggered when all objects are clustered in one unit (Algiriyage, Jayasena, & Dias, 2015). The dissimilarity measurement in this research is Ward's minimum variance method. Ward's method merges the pair of clusters with minimum between-cluster distance. In addition, the Ward' method clusters the objects based on multidimensional variance (Husson, Josse, & Pagès, 2010). When conducting the hierarchical clustering, a desired number of clusters (called *k*) will be identified by the algorithm showing in the dendrogram (Balcan, Liang, & Gupta, 2014).

K-means clustering

In contrast to hierarchical clustering, K-means clustering does not use hierarchies but aims to partition the numerical data into *k* clusters at once, using the nearest mean. As a typical partitional cluster algorithm, K-means is widely used in user segment, online profiling and behavioral targeting since it is simple to understand and implement (Xie, Jiang, Xie, & Gao, 2011) The process of K-mean clustering is indicated as following

(Qin, Guan, Wang, & Liu, 2015):

- Step 1: Initialize the cluster number (*K*)
- Step 2: Select the centroids of the clusters randomly
- Step 3: Compute the distance between the rest dataset
- Step 4: Update the center for each cluster

In the steps of k-means clustering, the cluster number has to be determined at first. Moreover, after randomly selecting the centroids, Euclidean distance is used to allocate the remaining data to the closest cluster (Zahra, Ghazanfar, Khalid, & Azam, 2015).

2.3.3. Multiple Correspondence Analysis

Correspondence analysis (CA) is defined as a statistical visualization method of the association between two level of categorical variables (Khangar & Kamalja, 2017). Instead of two level categorical variables, Multiple correspondence analysis (MCA) is used for multi-level of categorical variables. As the k-means clustering is applied to deal with numerical data, categorical data is not applicable directly in k-means clustering but needs to combine with MCA (Dam & de Velden, 2015; Mitsuhiro & Yadohisa, 2015). Thus, the categorical data is grouped with the MCA k-means clustering.

2.3.4. Adaption to this research

In this research, multiple online behaviors treated as categorical data are used to explore the students' online behavioral patterns. Hence MCA is applied to analysis the multiple categorical data. Before having the k-means clustering, agglomerative hierarchical clustering is conducted to find out the optimal cluster number k. Then the k-means algorithm is used to cluster the underlying behavioral patterns.

After having the clusters of behavioral patterns by the use of machine learning algorithms, the results of different clusters are explained based on the Hierarchy of Effects model in this research.

2.4. The Hierarchy of Effects model

As an extension of AIDA model (action, interest, desire and action), the Hierarchy of Effects model is developed by Lavidge and Steiner (1961) and is shown in figure 1. The model suggests there are six steps from viewing and seeking for information about the products/service to the final purchasing stage which is the conversion stage. Mokhtar (2016) states that the consumers could skip the steps during the process before reaching to the final conversion stage. This means that the consumers do not strictly follow the sequence of each step in the whole process.

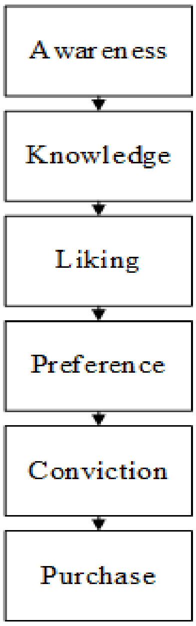


Figure 1. Hierarchy of effects Model (Karlsson, 2007)

When adapting to this research, the behavior of online application submission is regarded as the final conversion stage. Thus, the conversion funnel shows it as follows. Based on this model, the first stage (awareness) is recognizing the existence of master education of university of Twente on the website in cognitive level. The second stage (knowledge) is about knowing the core message delivered by the university about the programs offered in master study. Next stage (liking) is developing favorable attitude toward programs. Liking is not enough as the consumers may have favorable attitudes toward other similar programs offered by other universities. Thus, in the next step (preference), the potential customers have to prefer the master program in the university of Twente to the others. However, the preference does not necessarily transfer to the conversion in the view of this model. It means that the university has to persuade the potential consumers via the information on website that the program is the best selection to satisfy their expectations referring to conviction stage (Wijaya, 2012). Then it could move on to the last step (purchase) in which the users take actions to apply for the program. Online behaviors are conducted in each stage before reaching the final conversion.

3. Methodology

In this section, data description, measurement and analysis are outlined and described.

3.1. Dataset Description

The targeted group in this research are the students who have visited information about master studies and submitted the study application on the website of university of Twente from July 2016 to August 2017. The targeted students are divided into admitted students and non-admitted students as shown in figure 2. The reason about choosing the targeted group is that the university interests to know the online behavioral patterns of the students who submitted the application and searched master related information.

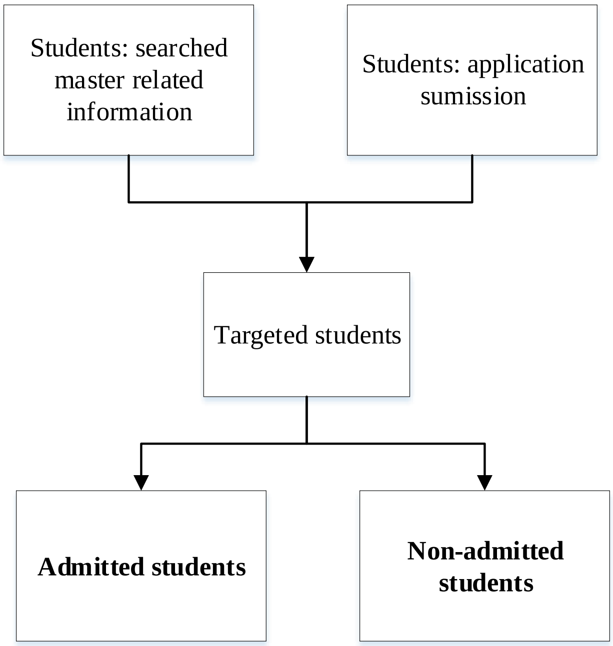


Figure 2. The targeted group

Two types of data extracted from the targeted students are used in this study. One type is the web data recording the online behaviors of the targeted students (both admitted and non-admitted students). The web data is extracted from Google Analytics as categorical data about whether they conducted the online behaviors or not. Another type is customer relationship management (CRM) data. CRM data have the static information of admitted students from 2016 to 2017, such as education level, faculty and program. Both types of data are offered and authorized by the Department Marketing and Communication (M&C) of the University of Twente. There are 2,840 targeted students which are formed by 528 admitted students and 2312 non-admitted students. The 2,840 targeted students have 9,040 categorical data points from the web data and CRM data.

The New European privacy regulation has come into force since 25 May 2018 which is called the General Data Protection (GDPR) (European Commission, 2018). The objective of this regulation is to protect the personal data of the natural persons, to reach the uniformity throughout the Europe and to promise the free movement of the personal data. According to the announcement of the university (University of Twente, 2018), university of Twente has informed its subjects including students, staff, prospective students, alumni and external staff about how and where data will be used. Thus, the data in this research are accordance with the GDPR.

3.2. Measurements

There are 10 types of student' online behaviors which are:

Table 2. Description of online behaviors

Online Behavior	**Description**
Check admission requirement	Check general admission related information
Education brochure request	Require to download the education brochure
Open day request	Register to an open day
PDF download	Download non-study related documents from the website, such as scholarship, new students information, and exchange students information,
Eligibility check	Have eligibility check for master study
Visit program webpage	Look for program related information
Visit master webpage	Look for master related information
Scholarship finder	Search scholarship related information
Question via web form	Ask online questions by filling in the web form
Check how to apply	Check the procedure of application

In addition, the conversion in research is the number of application submissions.

3.3. Data analysis

This sub-section outlines the steps in the data analysis process.

3.3.1. User profiling analysis process

The analysis process is formed by two parts. The first part is clustering analysis using a two-step MCA clustering model. The clustering analysis is conducted with the R software. In this part, the behavioral patterns of the targeted students (both admitted and non-admitted students) are discovered and categorized into different behavior profiles. Hence, the similar and different behavioral patterns are found out between the admitted and non-admitted students when applying for the university. Another part is discovering the demographic characteristics of the admitted students in each behavior profile, as there are CRM data about education level, faculty, and programs of the admitted students. It will be conducted with IBM Watson Analytics. Then the customer could be targeted with the comprehensive user profiles of the targeted students.

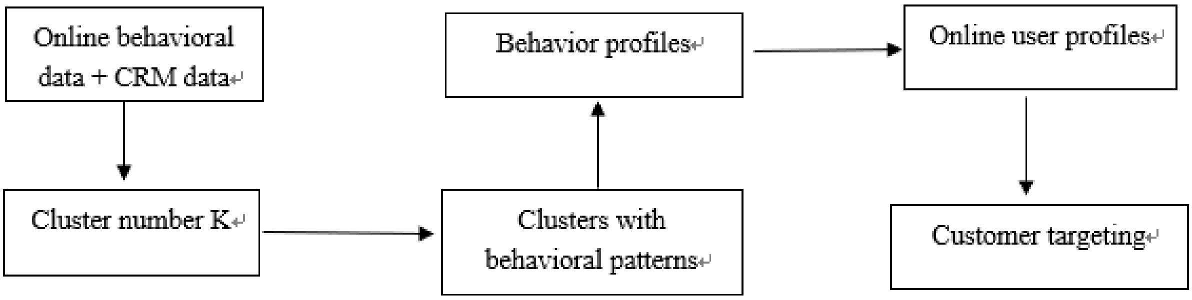


Figure 3. Data Analysis Process

3.3.2. Two-step MCA clustering model

A two-step MCA clustering model is conducted in this research. R software is applied with the program 'MCA', 'HCPC' and 'kmeans' in the package 'FactoMineR' and 'factoextra'. Before having the clustering analysis, the multiple correspondence analysis (MCA) is first used to analyze the categorical online behavioral data. Based on premise of MCA, hierarchical agglomerative clustering with ward distance will be utilized to determine the number of clusters. With the prior specified number of clusters, K-means could be operated to result in the clusters of online behavioral patterns. The process of the two-step MCA clustering model is shown in figure 4.

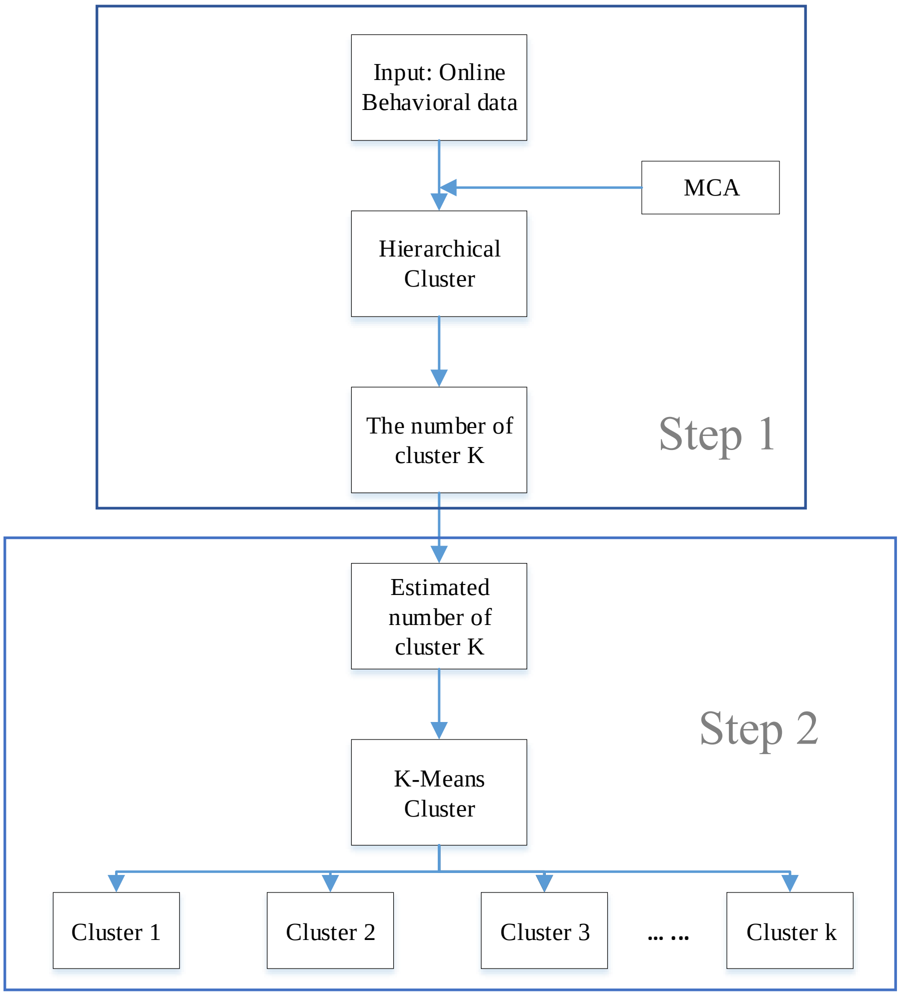


Figure 4. Two-step MCA clustering model

Step 1: Hierarchical Agglomerative Clustering algorithm

In hierarchical agglomerative clustering algorithm, the distance is computed based on total sum of squares within clusters (ward distance). First, each specific online activities (object) is considered as a cluster. The program will compress the raw input data into manageable sub clusters. Next, clustering will run to find and merge the most similar clusters in hierarchical structure. In the stage of cluster output, the number of clusters will be presented when satisfying the termination condition.

Step 2: K-means algorithm

With the known cluster number k, K-means algorithm will first initialize the center of the clusters. In the next stage, each online behavior will be allocated to the closest cluster center based on Euclidean distance. The new clusters will be renewed according to the assigned objects until the termination condition is reached. The output of K-means algorithm is the clusters of online behavioral patterns of the admitted and non-admitted targeted students.

3.3.3. Characteristics description

Characteristics of the admitted students in each behavior profile are formed based on the combination of CRM data and web data. The characteristics are ISO country code (the countries where the students performed online behaviors), source and medium, program, faculty and educational level. The prospective students could be targeted precisely with the characteristics. In this stage, IBM Watson analytics is utilized to offer a better visualization about the characteristics of the target group.

4. Results

4.1. General Data Description

There are 1,162,914 visitors recorded on the Google Analytics from1 July 2016 to 31 August 2017. However, they are unstructured with repeated entries. After filtering and cleaning the data, there are 17,363 students who searched master related information 1 July 2016 to 31 August 2017. In addition, 11 % of the students (2,312) have submitted the study application but not been admitted. Among the number of students who searched master related information, 528 students have been admitted occupying 3 % of the total number of students. The numbers and proportions of the targeted group can be seen in table 3 and figure 5.

Table 3. The number of students from July 2016 to August 2017

Total visitors (unstructured data)	1,162,914
Visitors who have searched master related information	17,363
Application submitted students	2,840
Application submitted students ---non-admitted	2,312
Application submitted students---admitted	528

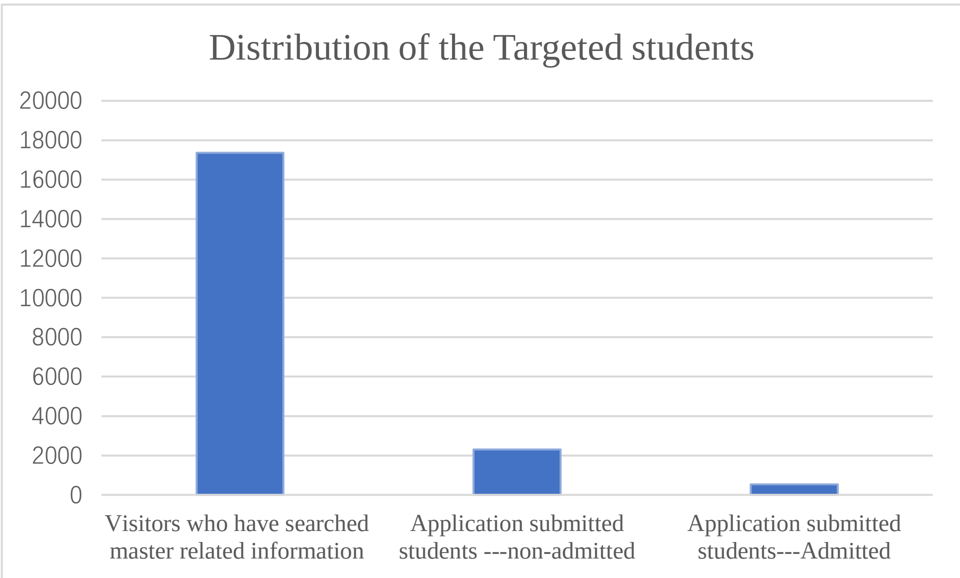


Figure 5. Distribution of targeted students

4.2. Clustering analysis---non-admitted students

4.2.1. Data description

From table 4, the students averagely have had two online actions through the website of the university before applying for the university. Six out of the ten online behaviors have been conducted by more than 20% of the total students. Among the students, one fourth of the students have done the online behaviors of PDF download as well as Eligibility check. Besides that, open day request, check admission requirement, education brochure request, and scholarship finder have been done by around 23% of the students. However, program webpage visit, check how to apply, questions via web forms, and master webpage visit have rarely been conducted by the students before applying for the study application.

Table 4. Online Behaviors for non-admitted students

Online Behaviors	Size	%	Mean	StdDev
Check admission requirement	518	22	0,22	0,42
Education brochure request	526	23	0,23	0,41
Open day request	537	23	0,23	0,42
PDF download	580	25	0,25	0,43
Eligibility check	580	25	0,25	0,44
Visit program webpage	49	2	0,06	0,23
Visit master webpage	11	1	0,02	0,11
Scholarship finder	499	2	0,2	0,41
Questions via web forms	35	2	0,05	0,22
Check how to apply	30	1	0,04	0,21

Note [1]: Binary Data

Note [2]: N=2312

4.2.2. Hierarchical clustering

Before having the hieratical agglomerative clustering, MCA analysis has been run in R with use of 'FactoMineR' and 'factoextra' package to transform the categorical data to continuous data. The result is shown in figure 6 and 7 where Ward's method is used as the basis to identify the number of clusters. In figure 7, it can be seen that there are some overlaps among different clusters. The reason is that some of the online behaviors were commonly performed by many the targeted students despite of different online behavioral patterns before submitting the application. The optimal number of clusters is 6 in the hieratical agglomerative clustering.

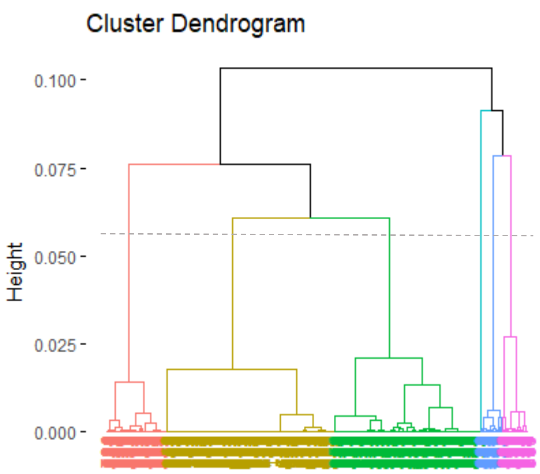


Figure 6. Hierarchical Agglomerative Clustering

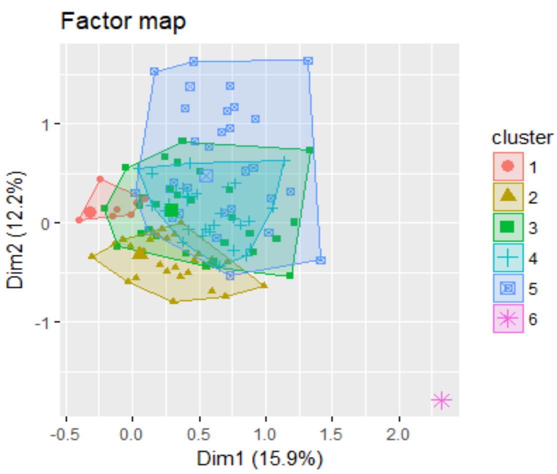


Figure 7. Factor map of Hierarchical Agglomerative

4.2.3. K-means clustering

The non-admitted students are divided into 6 clusters. It could be seen in table 5 that 40% of the students belong to cluster 2 which is the highest. The following cluster is cluster 4 including 35% of the total students. In addition, 14% of the students are in cluster 1 which is regarded as the third largest contributor. Whereas, cluster 3, 5 and 6 only consist of the rest 12% of the students where cluster 3 has the least number of students.

Table 5. Distribution of the k-means clusters

	N	**%**
Cluster 1	326	14
Cluster 2	925	40
Cluster 3	3	1
Cluster 4	799	35
Cluster 5	119	5
Cluster 6	140	6

The Distribution of the online behaviors in each cluster is indicated in table 6 and the specific statistics of each cluster is in appendix 2. In *cluster 1*, there is no online behaviors included before applying the university. It is interpreted that 14% of the students did not conduct any online behaviors before submitting their master application. In *cluster 2* where includes the most number of students, the behavioral pattern of the students is checking the admission requirement, having PDF download, and requiring for an open day before submitting the application. The centroid of cluster 2 is open day request. In *cluster 3,* the students prefer to have online contacts with the university before submitting the application. In *cluster 4*, the students' online behaviors are education scholarship finder, brochure request, and eligibility check. In this cluster, eligibility is the centroid. Different from the other clusters, behavioral pattern of *cluster 5* has visit program webpage and visit master webpage. In this cluster, the centroid is program webpage visit. In the last cluster, the students only check the admission requirement before applying.

Table 6. K-means online behavior distribution in each cluster for non-admitted students

	The Number of Clusters					
Online Behaviors	**1**	**2**	**3**	**4**	**5**	**6**
Check admission requirement	1%	62%	27%	10%	36%	75%
Education brochure request	0%	14%	34%	69%	32%	5%
Open day request	0%	50%	30%	3%	12%	2%
PDF download	1%	65%	27%	8%	26%	22%
Eligibility check	0%	42%	30%	43%	38%	4%
Visit program webpage	0%	1%	0%	0%	84%	0%
Visit master webpage	0%	0%	0%	0%	62%	0%
Scholarship finder	0%	37%	32%	42%	22%	13%
Questions via web forms	0%	0%	100%	0%	0%	0%
Check how to apply	0%	3%	50%	6%	0%	0%
Application Submission	100%	100%	100%	100%	100%	100%

4.3. Clustering analysis--- admitted students

The admitted students have different online behaviors from the rest. Thus, the online behaviors of 528 admitted students who searched master related information and submitted the application from 2016 to 2017 are explored in this section.

4.3.1. Data Description

Table 7. Online behaviors for admitted students

Online Behaviors	size	%	Mean	StdDev
Check admission requirement	125	24	0,237	0,426
Education brochure request	76	14	0,144	0,352
Open day request	285	54	0,541	0,499
PDF download	63	12	0,12	0,325
Eligibility check	137	26	0,26	0,439
Visit program webpage	427	81	0,81	0,392
Visit master webpage	145	28	0,275	0,447
Scholarship finder	202	38	0,383	0,487
Questions via web forms	32	6	0,061	0,239
Check how to apply	110	21	0,209	0,407

Note [1]: Binary Data

Note [2]: N=528

From table 7, the distribution of the online activities, during the process of application, four fifth of the students have browsed the program webpage to get more information about their preferred program before applying for the program. In addition, more than half of the students have requested for an open day during the process of application determination. Besides that, 38% of the students would like to seek for the information about scholarship before determining to submit the application. In addition, one fifth of the students check the procedures of application before applying.

4.3.2. Hierarchical clustering and elbow method

After transferring the categorical data with MCA, from figure 8, 5 clusters are identified as the optimal number of clusters in the hieratical agglomerative clustering.

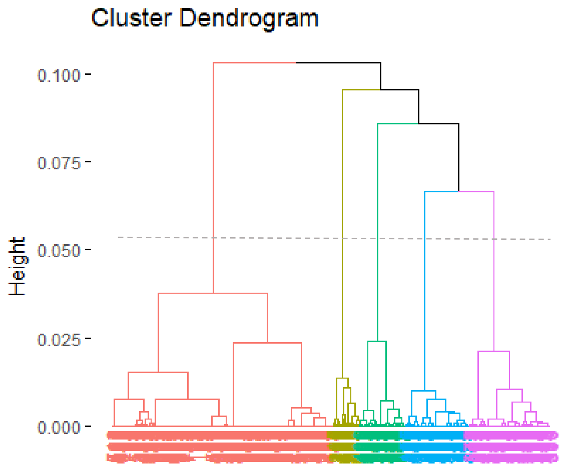


Figure 8. Hierarchical Agglomerative Clustering

4.3.3. K-means clustering

The distribution of the clusters is shown in table 8 where cluster 4 and cluster 2 are regarded as the top 2 clusters covering 33% and 23% of the admitted students respectively. Furthermore, cluster 1, cluster 3 and cluster 5 are averagely distributed by in k-means clustering.

Table 8. Distribution of the clusters

	N	%
Cluster 1	76	14
Cluster 2	120	23
Cluster 3	82	16
Cluster 4	175	33
Cluster 5	75	14

Specifically, in table 9, the number represents the value of each variable contributing to each cluster. It can be seen from the table that 6 online behaviors have large differences from different clusters. They are check admission requirement, education brochure request, open day request, eligibility check, visit program webpage, and scholarship finder. The differences represent that these variables are primary factors leading to different clusters.

In *cluster 1*, centroid is program webpage visit. The behavioral pattern of cluster 1 consists of program webpage visit, open day request, scholarship finder, admission requirement check, education brochure request, eligibility check and master webpage visit. In *cluster 2*, the central data point is the online behavior: check how to apply. The behavioral pattern of cluster 2 consists the online behaviors of checking how to apply, visiting program webpage, and finding scholarship. In *cluster 3*, scholarship finder is the central point. The behavioral pattern of cluster 3 consists of admission requirement check and scholarship finder. Although the online behaviors are also included in cluster 1, the centroids between these two clusters are different. In addition, the rest online behaviors in cluster 1 do not have a significant contribution to the behavioral pattern in cluster 3. In *cluster 4*, open day request is selected as the centroid in R software. the behavioral pattern of cluster 4 is open day request and program webpage visit. In *cluster* 5, the central seed is eligibility check. In addition, the behavioral pattern of cluster 5 is eligibility check, open day request, program webpage visit and scholarship finder.

Table 9. K-means Online behavior distribution in each cluster for admitted students

	The Number of Clusters				
Online Behaviors	**1**	**2**	**3**	**4**	**5**
Check admission requirement	86%	6%	66%	4%	0%
Education brochure request	50%	4%	11%	13%	17%
Open day request	66%	0%	6%	100%	72%
PDF download	24%	9%	10%	8%	15%
Eligibility check	62%	6%	22%	0%	100%
Visit program webpage	97%	98%	0%	92%	96%
Visit master webpage	67%	16%	15%	30%	36%
Scholarship finder	72%	22%	74%	29%	56%
Questions via web forms	7%	2%	10%	6%	9%
Check how to apply	36%	50%	4%	9%	7%
Application Submission	100%	100%	100%	100%	100%

4.4. Comparison analysis

In this sub-section, the comparison of behavioral patterns between non-admitted students and admitted students is outlined.

4.4.1. Clustering analysis

There is a significant difference between the admitted students and the non-admitted students in behavioral patterns. Firstly, for the non-admitted students, 14% of them did not show any online behaviors before having Osiris application. Besides that, the non-admitted students averagely have 2 online actions in each cluster. However, the admitted students averagely have 5 online behaviors when applying for the study which are more than the online behaviors of non-admitted students. In addition, for non-admitted students in cluster 2 which includes the largest number of students, they mainly prefer to check the general admission requirement and register for an open day to visit the university. However, for the admitted students in cluster 4 which includes the largest number of students, it shows that one third of the admitted students would like to look into the information about the programs which they are interested in by visiting the webpages of the available study programs besides checking the general admission requirement and requiring for an open day. Moreover, when looking into each specific online behavior, the differences are mainly reflected in open day request, visit program webpage, and check how to apply. The behavior of open day request is included in 3 out of 5 clusters of admitted students, but this behavior is only included in one cluster of non-admitted students. In addition, in the behavioral pattern of admitted students, they prefer to have a more comprehensive understanding about the information of the programs than the non-admitted students. It could be interpreted that the admitted students tend to gather more information about the university and the preferred program than the non-admitted student. Moreover, few non-admitted students choose to check how to apply before submitting the application as it is not included in any of the clusters.

4.4.2. Hierarchy of effects model in behavioral targeting

Based on the hierarchy of effect model, the behavior patterns in each cluster are explained in figure 9.

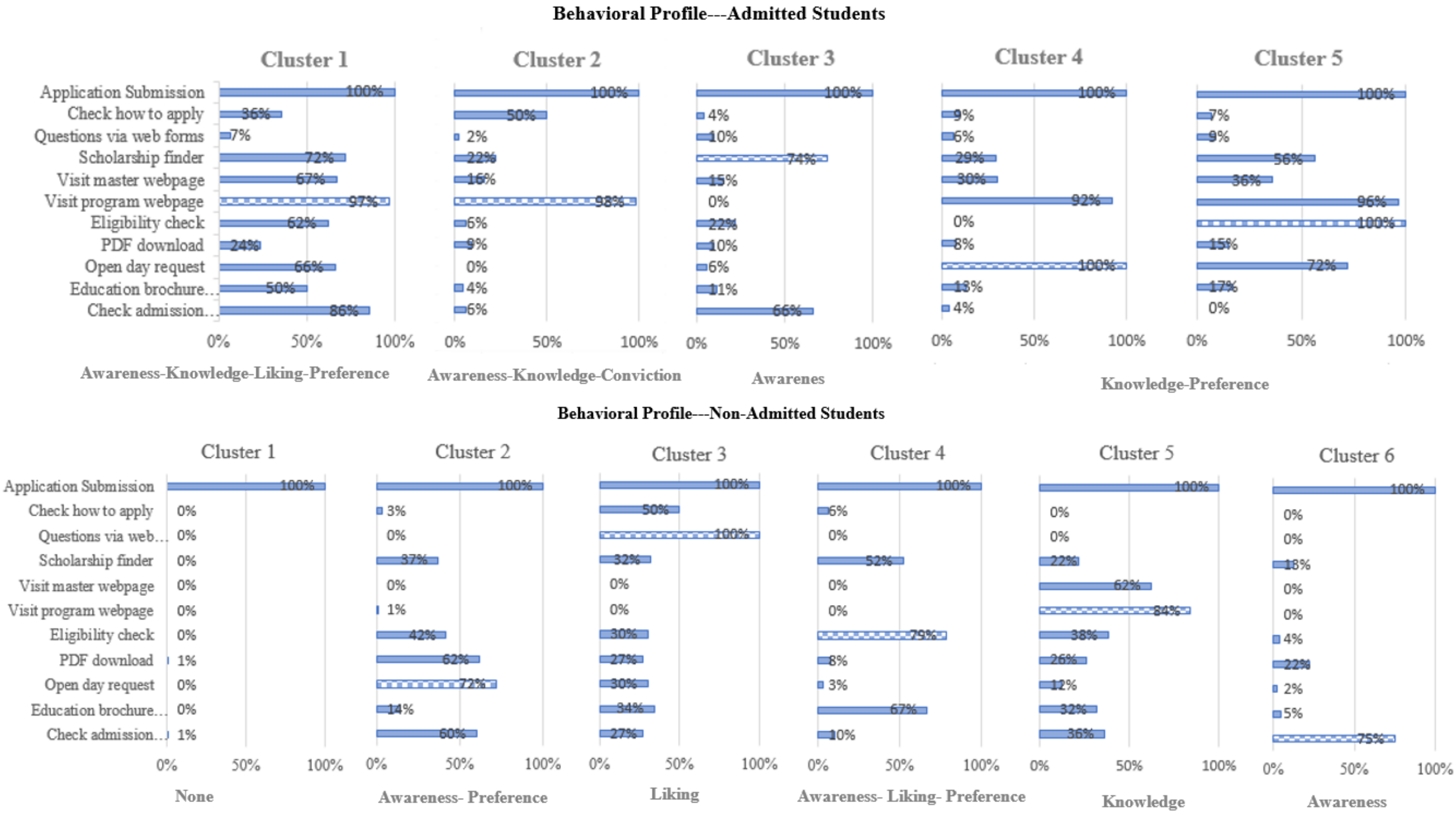


Figure 9. Behavioral Profile of admitted students and non-admitted students

For non-admitted students, there are five behavioral profiles that could explain the clusters which is also shown in figure 9 and 10.

<u>Behavioral profile 1: Awareness---Preference</u>

This behavioral profile contains 40% of the students which is the largest behavioral profile. In this profile, the students tend to check more general information about the university and prefer to see if they meet the basic admission requirement. Besides, with the basic impression about the university and program, they choose to register for an open day directly to get more information they need before deciding to apply for the university.

<u>Behavioral profile 2: Liking</u>

In this behavioral profile, the students would choose to submit the application when they found the program is interested to apply after having online contact with the university. However, only few of them follow this behavioral profile.

Behavioral profile 3: Awareness---liking---preference

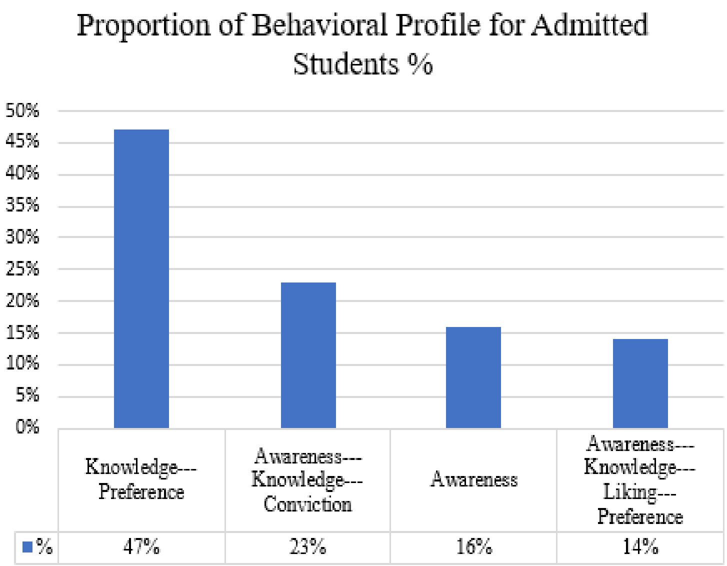
Different from behavioral profile 1, the students belonging to this behavioral profile require to download the education brochures about the programs they are interested in. It assists the student having an overview about programs they like before applying.

Behavioral profile 4: Knowledge

This profile represents that the only online behavior the students conducted before applying for the program was learning to get more knowledge about the programs. However, only 5 % of them have had these actions during the process of application.

Behavioral profile 5: Awareness

The students in this type of behavioral profile decided to submit the application as long as they aware they meet the basic admission requirement of the university. However, they did not learn and search more information about the university and the applied program. After all, awareness is the initial cognitive level of the in the Lavidge and Steiner hierarchy of effects model.



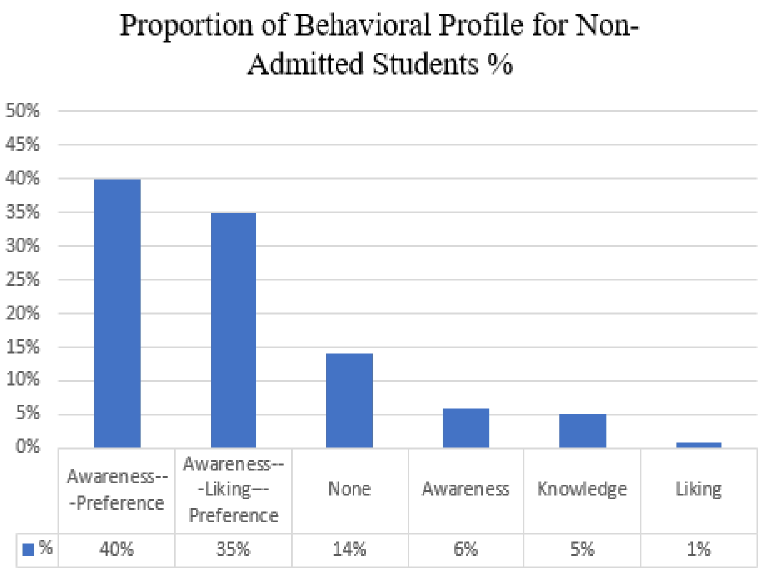


Figure 10. Behavioral Profile Distribution for Admitted Students and Non-admitted Students

For the admitted students, the five clusters are categorized into four types of behavioral profiles (see figure 9 and 10).

Behavioral profile 1: Awareness---knowledge---liking---preference

This profile includes 4 stages based on the Lavidge and Steiner hierarchy of effects model which means the students almost pass through all stages to reach the final stage of purchase. It means that in this type of behavior profile the students have had the awareness about the basic requirement and situation of the university. Besides, the students chose to acquire more knowledge about the programs available from the website of the university. By having the awareness and knowledge, the students have a good understanding about the university and the program in the university in cognitive level. In addition, the students are affected to have the interests in the programs. It means the students have had a positive attitude about their interested programs by downloading the brochure, having eligibility check and registering for an open day. With the behaviors in cognitive and affective level, the students have converted to the last stage that eventually submitting Osiris application. However, this behavioral profile contains the least number of grouped students.

Behavioral profile 2: Awareness-knowledge-conviction

In this behavioral profile, the students are interested in the scholarship information about the university. Besides finding the scholarship information, the admitted students prefer to have more knowledge about the program by vising the program webpages. In addition, they also checked how to apply to know the application procedure before eventually submitting the application.

Behavioral profile 3: Awareness

Similar to the behavioral profile 5 in clusters of non-admitted students, this profile only includes awareness. It represents that the grouped students considered the cognitive dimension as the key to conversion. However, what is different from the behavioral profile of non-admitted students is that the admitted students have more online behaviors than the non-admitted students in the stage of awareness. The admitted students preferred to check the general admission requirement and find the scholarship information. Once the students found they met the requirement and could afford the cost, they would choose to apply for the university.

Behavioral profile 4: knowledge---preference

Cluster 4 and cluster 5 are categorized in behavioral profile 4 which includes almost half of the admitted students. Comparing with other behavioral profiles, the students in this profile consider about the quality of the program more as the key when deciding to have the application. The reason is that they would like to have a comprehensive knowledge about their interested programs by visiting the online program webpages. In addition, in the preference stage, by registering to go for the open day and having eligibility check, the students found they would like to apply for the program and at the same time they are qualified to apply for the program as well. Thus, the applications were submitted eventually.

To summarize, taking the behavioral profiles of both admitted and non-admitted students into consideration, awareness and preference are more related to the application submission. However, the admitted students are averagely involved in more stages before reaching the last purchase step. Besides, the stage of knowledge becomes the key difference between admitted students and non-students when submitting the application.

4.5. User profile description

When it comes to user profile description. the group of admitted students are the "good samples" as they are admitted after submitting Osiris application when searching master related information during the process of application. Thus, it is worth to explore the characteristics of these group of students in order to assist the university in targeting prospective students. Combined with the available customer relationship management (CRM) data, several available indicators are considered in this research including education level, faculty, sources/medium and interest country ISO code.

Educational level indicates if the students are admitted as bachelor or master students. Faculty means the faculty of the students who belong to. Sources/medium shows the online channels that the students used to access the website of university. Interest country ISO code means that the location of IP address where the students had the online behaviors, but it does not necessarily stand for the nationality of the students. In general, it could be resulted that the distributions of the CRM indicators have no significant differences between each cluster. It means that the characteristics of the students are similar to each other between clusters.

Table 10. Distribution of Education level in each cluster

	Cluster number						
Education level	1	2	3	4	5	**Total**	%
Bachelor	43	60	43	96	32	274	52
Master	33	60	39	79	43	254	48

The distribution of the education level (table 10) shows that instead of master students, more than half of the students who searched master information and submitted the application are actually bachelor students. Following by the bachelor students, the rest half of the students searched the master related information is mainly for applying master program in the university.

Table 11. Distribution of Faculty in each cluster

	Cluster number						
Faculty	**1**	**2**	**3**	**4**	**5**	**Total**	%
Engineering Technology (ET)	18	25	16	37	15	111	21
Electrical Engineering, Mathematics and Computer Science (EWI)	20	21	10	31	9	91	17
Science and Technology (TNW)	10	21	22	41	14	108	20
Behavioral, Management&Social Sciences (BMS)	26	51	34	62	33	206	39
N.A.	2	2	0	4	4	12	2

There are four faculties as shown in table 11 where more students (39 %) are from BMS. The distributions are the same for all behavioral patterns. While the faculty of EWI has the least number of students.

Table 12. Distribution of main online behavior locations in each cluster

	Cluster number						
Interest country ISO code	**1**	**2**	**3**	**4**	**5**	Total	%
India	1	2	5	6	4	18	3
Germany	2	23	8	15	4	52	10

The Netherlands	62	75	57	134	60	388	73

Note: the full picture of locations is in appendix 4

When it comes to the interest country where the students have online behaviors, it could be concluded that the majority of admitted students have conducted their online behaviors during the process of application in the Netherlands. Following by the Netherlands, Germany and India are the second and third favorable countries where the students have online behaviors. Thus, the students who searched the master related information and submitted the application are mainly from these three countries accounting for 86 % as shown in table 12. The rest 14 % of the students are located in other countries (the complete distribution of Interest country ISO code is in appendix 4)

Table 13. Distribution of source/medium in each cluster

	Cluster number						
Source/medium	1	2	3	4	5	**Total**	%
outlook/live.com/referral	3	3	1	5	8	20	4
google/cpc	6	10	3	10	5	34	6
google/organic	57	77	54	113	53	354	67

Note: the full picture of source is in appendix 4

There are diverse online channels to access the webpages of the university and have their online behaviors, such as, email, search engine, Studielink, Facebook, direct entry, and other educational website (see complete distribution table in appendix 4). However, as extracted in table 14, most of the students choose to access the webpages of the university to have their online behavior from google accounting for 67 percent. It indicates that although there are variable online media, google is still the first choice for the students when accessing to the website before having online behaviors

Table14. Main Targeted student distribution in each cluster

				Cluster number						
Faculty	**Source/Medium**	**Location**	**Education Level**	**1**	**2**	**3**	**4**	**5**	**Total**	%
BMS	google	The Netherlands	BSc	14	11	8	22	11	66	13
BMS	google	The Netherlands	MSc	5	4	8	10	13	40	8
EWI	google	The Netherlands	BSc	10	5	6	12	6	39	7
TNW	google	The Netherlands	BSc	3	7	7	13	3	33	6
EWI	google	The Netherlands	BSc	7	7	3	10	2	29	5
TNW	google	The Netherlands	MSc	1	7	4	7	2	21	4
BMS	google	Germany	BSc	0	5	3	11	1	20	4
EWI	google	The Netherlands	MSc	3	7	2	2	2	16	3
EWI	google	The Netherlands	MSc	4	4	2	4	1	15	3
BMS	google	Germany	MSc	1	7	2	1	1	12	2
TNW	google	The Netherlands	BSc	1	1	1	6	3	12	2

Note: the full picture of distribution is in appendix 4

When integrated with all CRM indicators (the completed integrated table is in appendix 4), it could be interpreted that the students who searched master related information and submitted the application are mainly admitted as the bachelor students from the faculty of BMS who use google as organic search engine to have online application behaviors on the website of the university in the Netherlands

5. Conclusion

5.1. Key findings

The key findings in this research focus on the *behavioral profiles* of the students who searched master related information and submitted the application from 2016 to 2017, the *differences* between admitted students and non-admitted students in the targeted group, and the *characteristics* of the admitted students.

Both admitted and non-admitted students presented online behaviors by checking general admission requirement, finding scholarship information, registering for an open day and having eligibility check in their behavioral profiles before having Osiris applications. In other words, once the students have the awareness about the basic study requirement and scholarship information of the university, and have the interests to requiring for an open day and eligibility check, they are more likely to submit the application. In addition, the second key finding is that the students who have the online behaviors of visiting the program webpages in their behavior profiles have more likelihood to be admitted. Moreover, about the user profile description, the students who searched master related information and submitted the application are mainly admitted as the bachelor students from the faculty of BMS who use google as organic search engine to have online application behaviors on the website of the university in the Netherlands.

5.2. Implication

5.2.1. Theorical implication

This research has developed an analysis approach which shows the process of behavioral targeting including data extraction, clustering analysis, behavioral profile, and customer characteristics profile. In clustering analysis, a two-step MCA clustering model is developed to group the online behavioral data into different clusters. Besides that, this model is particular used to deal with the categorical data with multiple dimensions. In addition, the research extends the scope of the Lavidge and Steiner hierarchy of effects model to assist the university in students' recruitments.

5.2.2. Practical implication

The behavior profiles could assist the university placing the online advertising on the webpages where the students browser, visit, click and download frequently. Besides, the research suggests that the university could make marketing communication strategy to encourage the prospective students to develop the awareness and preference of the university and the programs by having relevant online behaviors. In addition, as the admitted students have more knowledge about the program by visiting the program webpage, the university is suggested to promote the prospective students to have a relative comprehensive knowledge about the programs before applying to the university in order to increase the admitted rate of the university. When it comes to the characteristics of the students, as more admitted students in the targeted group are bachelor students from the faculty of BMS who use google as organic search engine to have online application behaviors on the website of the university in the Netherlands, the characteristics could be contributed to have precise targeting about prospective students for the university.

6. Discussion

The objective of this research is to explore online behavioral profiles of the students who searched master related online information and submitted online university application from 2016 to 2017 and to develop a behavioral targeting analysis approach with the use of machine learning

First of all, the analysis approach has been developed (see section 3.3 and figure 3) in this research. In the approach, online behavioral data initially needs to be extracted and cleaned. Next, the cleaned behavioral data would be run in cluster analysis. The cluster analysis follows a two-step clustering model with hierarchical clustering and non-hierarchical clustering (k-means). In addition, if the online behavioral data are categorical data with more than two variables, the two-step clustering model has to be operated in MCA. After getting the clusters, the behavioral profile could be developed based on the hierarchy of effect models. Next, combining with the CRM data, the characteristics of the customers in tach type of behavioral profile could be explored to assist in customer targeting.

Based on the results, the students who passed through the stage of awareness and preference have a significant correlation with the application submission which is the conversion in this research. According to Lavidge and Stener (1961), awareness and preference are significant contributors for the final conversion. Thus, the result is in line with the Lavidge and Steiner hierarchy of effects model. Specifically, the students with the behavioral pattern of checking admission requirement, requiring open day, and having eligibility check are more correlated with application submission.

However, comparing with the behavioral profiles between admitted students and non-admitted students, the admitted students have more number of online behaviors than the non-admitted student before submitting the application. In addition, the majority of admitted students have ‘visit program webpage’ in their behavioral patterns before applying to the study. However, this online behavior is not obvious for non-admitted students. Based on hierarchy of effects model, the users would have a rational understanding when having more knowledge about the interested products or service. Therefore, it can be resulted that the students who conducted the online behavior to acquire more comprehensive knowledge about the program have more likelihood to be admitted. The possible reason behind could be that the students could think more about if they really match the program by visiting the program webpage. The rational thinking would assist the students to be more objective when deciding to apply for the program which increases the successful rate to be admitted.

When it comes to the characteristics the admitted students, there is an interesting finding that more than half of the targeted students who searched master related information actually applied for bachelor study and admitted as bachelor students. This finding implies that after graduating from the high school many students have a related long-term higher education planning that is staying in the same university for their bachelor and master studies. Thus, even the students actually applied for bachelor program in the university, they would like to take actions to know more about the information about master in the university before making decisions to submit the application. In addition, the distributions of the characteristics of the students do not have significant differences between different behavioral profiles. Integrating with the other CRM indicators, the students who searched master related information and submitted the application are mainly admitted as the *bachelor students* from the *faculty of BMS* who use *google* as organic search engine to have online application behaviors on the website of the university in *the Netherland*s.

7. Limitation and Future research

There are two main limitations in this research. First, online behavioral data were only available from July 2016 to August 2017, thus the online behaviors before July 2016 were not included in the database. In addition, although the methods about the behavioral targeting could be generalized to other relevant research, the specific results about the behavior profiles could be not generalized as the clustering analysis has limitation about generalization. In this case, the results of this research could not be directly copied to recruit perspective students in other universities.

About the future research, first, this research has explored the relationship between online behavioral patterns and application submission of the targeted students. We could conclude that some online behaviors such as check admission requirement and open day request are more related to the application submission for both admitted and non-admitted students, but it has not been proved that those online behaviors lead to the application submission in this research. Therefore, in the future research, it is worth to investigate if there is a causal relationship between the online behaviors and application submissions. In addition, the future research could also explore how did the students who have applied the university create the awareness at the beginning. In other words, which online behavioral patterns drive the prospective students from unawareness to awareness is interested to investigate that could assist the university developing a full picture of the customer journey.

8. References

Ahmed, A., Low, Y., Aly, M., Josifovski, V., & Smola, A. J. (2011). Scalable Distributed Inference of Dynamic User Interests for Behavioral Targeting. *ACM SIHKDD International Conference on Knowledge Discovery and Data Mining*, 114-122.

Algiriyage , N., Jayasena , S., & Dias, G. (2015). Web user profiling using hierarchical clustering with improved similarity measure. *Moratuwa Engineering Research Conference (MERCon)* , 295-300. doi:10.1109/MERCon.2015.7112362

Al-Qurishi, M., Alhuzami, S., AlRubaian, M., Hossain, S. M., Alamri, A., & Rahman, A. M. (2018). User profiling for big social media data using standing ovation model. *Multimed Tools Appl, 77*, 11179–11201. doi:https://doi.org/10.1007/s11042-017-5402-6

Amoretti, M., Belli, L., & Zanichelli, F. (2017). UTravel: Smart Mobility with a Novel User Profiling and Recommendation Approach. *Pervasive and Mobile Computing, 38*, 474-489. doi:http://dx.doi.org/10.1016/j.pmcj.2016.08.008

Arzubov, M., Shakhovska, N., & Lipinski, P. (2017). Analyzing ways of building user profile based on web surf history. *12th International Scientific and Technical Conference on Computer Sciences and Information Technologies (CSIT), 1*, 377-380. Retrieved from https://ieeexplore.ieee.org/stamp/stamp.jsp?arnumber=8098809

Bagherjeiran, A., Hatch, A., & Ratnaparkhi , A. (2010). Ranking for the Conversion Funnel. *Proceedings of the 33rd international ACM SIGIR conference on Research and development in information retrieval* , 146-153.

Balcan, M.-F., Liang, Y., & Gupta, P. (2014). Robust Hierarchical Clustering. *Journal of Machine Learning Research,15*, 4011-4051. Retrieved from http://www.jmlr.org/papers/volume15/balcan14a/balcan14a.pdf

Bang, J., Cho, Y., & Kim, M. (2015). Getting Business Insights through Clustering Online Behaviors. *Modelling and Simulation in Engineering*, 1-8. doi:10.1155/2015/914314

Barry , T. E., & Howard, D. J. (1990). A review and critique . of the hierarchy of effects in advertising. *International Journal of Advertising, 9(2)*, 121-135.

Barry, E. T. (1987). The Development of the Hierarchy of Effects:. *Current issues and Research in Advertising,10(1-2)*, 251-295.

Barry, E. T., & Howard, J. D. (1990). A review and critique of the hierarchy of effects in advertising. *International Journal of Advertising, 9(2)*, 121-135.

Beales, H. (2010). The Value of Behavioral Targeting. *Network Advertising Initiative*, 1-25.

Beane, T., & Ennis , D. (1987). Market Segmentation: A Review. *European Journal of Marketing, 21(5)*, 20-42.

Boerman, S. C., Kruikemeier, S., & Borgesius, F. J. (2017). Online Behavioral Advertising: A Literature Review and Research Agenda. *Journal of Advertising*, 1-14. doi:10.1080/00913367.2017.1339368

Boratto, L., Carta, S., Fenu, G., & Saia, R. (2016). An enhanced model for behavioral targeting in online advertising. *Knowledge-Based Systems,108*, 5-14.

Borgesius, Z. (2015). Improving privacy protection in the area of behavioural targeting. 28.

Breznitz , D., & Palermo , V. (2013). A strategic advantage with behavioral targeting? How can (and what) firms benefit from personal data-based online marketing strategies. *35th DRUID Celebration Conference*.

Brownlee, J. (2016). *Machine Learning Algorithms: Discover how they work and implement Them From Scratch.* Retrieved from https://machinelearningmastery.com/supervised-and-unsupervised-machine-learning-algorithms/

Çağil, G., & Erdem, M. B. (2012). An intelligent simulation model of online consumer behavior. *Journal of Intelligent Manufacturing, 23(4)*, 1015-1022. doi: 10.1007/s10845-010-0439-7

Caudill , E. M., & Murphy, P. E. (2012). Consumer Online Privacy: Legal and Ethical Issues. *In Security and Privacy (SP), 2012 IEEE Symposium*, 413-427. Retrieved from https://www.jstor.org/stable/pdf/41060701.pdf

Chen, J., & Stallaert, J. (2014). An economic analysis of online advertising using behavioral targeting. *MIS Quarterly,38(2)*, 429-450.

Chung , S., & Park , C. (2009). Online Shopping Behavior Model: A Literature Review and Proposed Model . *International Conference on Advanced Communication Technology, 3*, 2276-2282.

Cufoglu , A. (2014). User Profiling-A Short Review. *International Journal of Computer Applications, 108(3)*, 1-9. Retrieved from https://pdfs.semanticscholar.org/eecb/f9358916a8e7db20511c611eaceaac554417.pdf

Dam, J.-W. v., & de Velden, M. v. (2015). online profiling and clustering of facebook users. *Decision Support Systems,70*, 60-72.

de Amorim , R. C., & Komisarczuk, P. (2014). Chapter 18: Partitional Clustering of Malware Using K-Means. In C. Blackwell, & H. Zhu, *Cyberpatterns: Unifying design patterns with security and attack patterns* (pp. 223-233). Springer International Publishing.

Deane, J. K., & Meuer, T. (2011). A longitudinal analysis of web surf history to maximise the effectiveness of behaviourial targeting techniques. *Int. J. Electronic Marketing and Retailing, 4*, 117-128.

Debatin, D., & Lovejoy , J. P. (2009). Facebook and online privacy: Attitudes, behaviors, and unintended consequences. *Journal of computer-mediated communication,15(1)*, 83-108. Retrieved from https://watermark.silverchair.com/jjcmcom0083.pdf?token=AQECAHi208BE49Ooan9kkhW_Ercy7Dm3ZL_9Cf3qfKAc485ysgAAAcQwggHABgkqhkiG9w0BBwagggGxMIIBrQIBADCCAaYGCSqGSIb3DQEHATAeBglghkgBZQMEAS4wEQQMw1WIMOE-7j3brH9OAgEQgIIBd4Ba1mGwAi2kkUhtMhm10Njs_ilhBmr4H98rIJCrd8J

Dolnicar, S. (2008). Market Segmentation in Tourism. In A. A. Woodside, & D. Martin, *Toursim management, Analysis. Behaviour, and Strategy* (pp. 129-150). Wallingford, Oxfordshire.

European Commission. (2018, July 06). *Data protection in the EU*. Retrieved from European Commission: https://ec.europa.eu/info/law/law-topic/data-protection/data-protection-eu_en

Fan, X., Chow, K.-P., & Xu, F. (2016). Web User Profiling Based on Browsing Behavior Analysis. *IFIP International Conference on Digital Forensics*, 57-71. doi:https://hal.inria.fr/hal-01393760/document

Gentleman , R., & Carey, V. J. (2008). Unsupervised Machine Learning. In F. Hahne, W. Huber , R. Gentleman, & S. Falcon, *Bioconductor Case Studies* (p. 137). New York: Springer Science+Business Media. doi:10.1007/978-0-387-77240-0

Germanakos, P., Tsianos, N., Lekkas, Z., Mourlas, C., & Samaras, G. (2009). Realizing comprehensive user profile as the core element of adaptive and personalized communication environments and systems. *The Computer Journal, 52(7)*, 749-770. Retrieved from https://pdfs.semanticscholar.org/2fc1/7c6fda003bc70e705febd21247cfa8fafcdf.pdf

Ghatak , A. (2017). *Machine learning with R.* Singapore: Springer Nature . Retrieved from https://link.springer.com/content/pdf/10.1007%2F978-981-10-6808-9.pdf

Ghirvu, A. I. (2013). THE AIDA MODEL FOR ADVERGAMES. *The USV Annals of Economics and Public Administration ,13(17)*, 90-98. Retrieved from http://www.seap.usv.ro/annals/ojs/index.php/annals/article/viewArticle/497

Godoy, D., & Amandi, A. (2005). User profiling in personal information agents: a survey. *The Knowledge Engineering Review, 20(4)*, 329-361. Retrieved from https://www.cambridge.org/core/services/aop-cambridge-core/content/view/383527ACC2A6C40746AD7FE48E7E8B1D/S0269888906000397a.pdf/div-class-title-user-profiling-in-personal-information-agents-a-survey-div.pdf

Goldfarb , A., & Tucker , C. E. (2011). Privacy Regulation and Online Advertising. *Management Science,57(1)*, 57–71. Retrieved from https://pubsonline.informs.org/doi/pdf/10.1287/mnsc.1100.1246

Gong, X., Guo, X., Zhang, R., He, X., & Zhou, A. (2013). Search Behavior Based Latent Semantic User Segmentation for Advertising Targeting, 13. *IEEE International Conference on Data Mining*, 211-220. doi:10.1109/ICDM.2013.62

Goyat , S. (2011). The basis of market segmentation: a critical review of literature. *European Journal of Business and Management, 3(9)*, 45-55.

Gupta, P., Sharma, A., & Jindal, R. (2016). Scalable machine-learning algorithms for big data analytics: a comprehensive review. *WIREs at JSM, 6(6)*, 194-214. doi:https://doi.org/10.1002/widm.1194

Hair, J. F., Black, W. C., Babin, B. J., & Anderson, R. E. (2013). 8. Cluster Analysis. In J. F. Hair, W. C. Black, B. J. Babin, & R. E. Anderson, *Multivariate Data Analysis: Pearson New International Edition* (pp. 415-474). Pearson Education Limited.

Hamka, F., Bouwman, H., de Reuver, M., & Kroesen, M. (2014). Mobile customer segmentation based on smartphone measurement. *Telematics and Informatics, 31*, 220-227. Retrieved from https://ac.els-cdn.com/S0736585313000440/1-s2.0-S0736585313000440-main.pdf?_tid=8235846a-e80e-42b4-b75b-3863df094081&acdnat=1531814448_0b323a26a202d78e8d73f57cbdca2ac3

Hassan, A., Jones, R., & Klinkner, K. L. (2010). Beyond DCG: User Behavior as a Predictor of a Successful Search. *WSDM 2010 - Proceedings of the 3rd ACM International Conference on Web Search and Data Mining* , 221-230.

Hassana, S., Nadzimb, S. Z., & Shiratuddin, N. (2015). Strategic Use of Social Media for Small Business Based on the AIDA Model. *Procedia - Social and Behavioral Sciences. 172*, 262 – 269. Retrieved from https://ac.els-cdn.com/S1877042815004000/1-s2.0-S1877042815004000-main.pdf?_tid=62333ef5-c0c6-4d58-9606-148ce080df5c&acdnat=1529317773_6d3b6bdd1dc0a4cf89f2abef60c3faac

Hawalah, A., & Faslia, M. (2015). Dynamic user profiles for web personalisation. *Expert Systems with Applications,42,(5)*, 2547-2569. doi:https://doi.org/10.1016/j.eswa.2014.10.032

Heath, R., & Feldwick, P. (2007). Fifty years using the wrong model of advertising. *International Journal of Market Research,5(1)*, 29-59.

Hruschka, H., & Natter , M. (1999). Comparing performance of feedforward neural nets and K-means for cluster-based market segmentation. *European Journal of Operational Research ,114*, 346-353. Retrieved from https://ac.els-cdn.com/S0377221798001702/1-s2.0-S0377221798001702-main.pdf?_tid=a4322244-9cfc-4cbb-9f6e-e0883006ea5d&acdnat=1531820086_73465b6cc046768637db00a34261d5b6

Husson, F., Josse, J., & Pagès, J. (2010). Principal component methods - hierarchicalclustering - partitional clustering: why would we need to choose for visualizing data? *Applied Mathematics Department*, 1-17. Retrieved from http://factominer.free.fr/more/HCPC_husson_josse.pdf

Hutter, K., Hautz, J., Dennhardt, S., & Fu¨ller, J. (2013). The impact of user interactions in social media on brand awareness and purchase intention: the case of MINI on Facebook. *ournal of Product & Brand Management,22(5-6)*, 342-351.

Imenda, S. N., Kongolo, M., & Gewal, A. S. (2004). Factors Underlying Technikon and University Enrolment Trends in South Africa. *Educational Management Administration & Leadership, 32(2)*, 195–215. doi: 10.1177/1741143204041884

Jain, P., Garibaldi, J. M., & Hirst, J. D. (2009). Supervised machine learning algorithms for protein structure classification. *Computational Biology and Chemistry, 33(3)*, 216-223. doi:https://doi.org/10.1016/j.compbiolchem.2009.04.004

Kanoje, S., Girase, S., & Mukhopadhyay , D. (2014). User Profiling Trends, Techniques and Applications. *International Journal of Advance Foundation and Research in Computer,1(1)*, 1-6. Retrieved from https://arxiv.org/ftp/arxiv/papers/1503/1503.07474.pdf

Karlsson, L. (2007). Advertising Theories and Models ñ how well can these be transferred from text into reality? *University of Halmstad* .

Khanchana , R., & Punithavalli, M. (2011). Web Usage Mining for Predicting Users' Browsing Behaviors by using FPCM Clustering. *International Journal of Engineering and Technology, 3(5)*, 491-496.

Khanchana, R., & Punithavalli, M. (2013). A Web Usage Mining Approach Based On New Technique In Web Path Recommendation Systems . *nternational Journal of Engineering Research & Technology, 2(1)*, 1-6.

Khangar , N. V., & Kamalja, K. K. (2017). Multiple Correspondence Analysis and its applications. *Electronic Journal of Applied Statistical Analysis , 10(2)*, 432-462. doi:10.1285/i20705948v10n2p432

Kotler, P., Armstrong, G., Wong, V., & Saunders, J. (2008). Chapter 9 Segmentation and Positioning. In P. Kotler, G. Armstrong, V. Wong, J. Saunders, & 05 (Ed.), *Principles of marketing* (p. 446). New Jersey, United States of America: Pearson Education. Retrieved from https://books.google.nl/books?id=6T2R0_ESU5AC&printsec=frontcover&source=gbs_ge_summary_r&cad=0#v=onepage&q&f=false

Krishnan , G. S., & Kamath , S. S. (2017). Dynamic and temporal user profiling for personalized recommenders using heterogeneous data sources. *8th International Conference on Computing, Communication and Networking Technologies (ICCCNT)*, 1-7. doi:10.1109/ICCCNT.2017.8203963

Kumar, A., & Ashraf, M. (2015). Personalized Web Search Engine using Dynamic User Profile and Clustering Techniques. *2nd International Conference on Computing for Sustainable Global Development (INDIACom)* , 2105-2108. Retrieved from https://ieeexplore.ieee.org/document/7100610/

Lavidge, R. J., & Steiner, G. A. (1961). A model of predictive measurement of advertising effectiveness. *Journal of Marketing, 25(6)*, 59-62. doi:10.2307/1248516

Lu, X., Zhao, X., & Xue, L. (2016). Is Combining Contextual and Behavioral Targeting Strategies Effective in Online Advertising? *ACM Transactions on Management Information System,7*, 1-20. doi:10.1145/2883816

M , R. V., Thottungal, A., & Nizar , A. (2016). An Enhanced Model for Behavioral Targeting in Online Advertising. *Proceedings of the 2016 International Conference on Data Science and Engineering*.

Manchanda , P., Dubé , J.-P., Goh , K. Y., & Chintagunta , P. K. (2006). The Effect of Banner Advertising on Internet Purchasing. *Journal of Marketing Research, 43(1)*, 98-108. doi:https://doi.org/10.1509/jmkr.43.1.98

Martín-Bautista, M. J., Kraft, D. H., Vila, M. A., Chen, J., & Cruz, J. (2002). User profiles and fuzzy logic for web retrieval issues. *Soft Computing,6(5)*, 365-372. doi:https://doi.org/10.1007/s00500-002-0190-x

Mingoti, S. A., & Lima, J. O. (2006). Comparing SOM neural network with Fuzzy c-means, K-means and traditional hierarchical clustering algorithms. *European Journal of Operational Research,174(3)*, 1742-1759.

Mitsuhiro, M., & Yadohisa, H. (2015). Reduced k-means clustering with MCA in a low-dimensional space. *Computational Statistics,30(2)*, 463–475. doi:https://doi.org/10.1007/s00180-014-0544-8

Mokhtar, A. (2016). A framework for Islamic advertising: Using Lavidge and Steiner’s hierarchy of effects model. *Intellectual Discourse, 24(2)*, 273-294.

Mustafal, N., Ibrahim, A. O., Ahmed, A., & Abdullah, A. (2017). Collaborative Filtering: Techniques and Applications. *2017 International Conference on Communication, Control, Computing and Electronics Engineering (ICCCCEE*, 1-6. Retrieved from https://ieeexplore.ieee.org/abstract/document/7867668/

Pandey, S., Aly, M., Bagherjeiran, A., Hatch, A., Ciccolo, P., Ratnaparkhi, A., & Zinkevich, M. (2011). Learning to Target: What Works for Behaviroal Targeting. *Informational Conference on Information and Knowledge Management, 20*, 1805-1814. doi:10.1145/2063576.2063837

Pandey, S., Aly, M., Josifovski , V., & Punera , K. (2013). Towards a Robust Modeling of Temporal Interest Change Patterns for Behavioral Targeting. *WWW 2013 - Proceedings of the 22nd International Conference on World Wide Web*, 71-81.

Poo, D., Chng, B., & Goh, J.-M. (2003). A hybrid approach for user profiling. *Proceedings of the 36th Annual Hawaii International Conference on*, 1-9. Retrieved from https://www.computer.org/csdl/proceedings/hicss/2003/1874/04/187440103b.pdf

Portugal, I., Alencar, P., & Cowan, D. (2018). The use of machine learning algorithms in recommender systems: A systematic review. *Expert Systems With Applications, 97*, 205-227. doi:https://doi.org/10.1016/j.eswa.2017.12.020

Qin, T., Guan, X., Wang, C., & Liu, Z. (2015). MUCM: Multilevel User Cluster Mining Based on Behavior Profiles for Network Monitoring. *IEEE Systems Journal,9(4)*, 1322-1333.

doi:10.1109/JSYST.2014.2350019

Qiu, J., Wu, Q., Wu, Y., & Feng, S. (2016). A survey of machine learning for big data processing. *EURASIP Journal on Advances*, 67. doi:10.1186/s13634-016-0355-x

Sadeghi, T., Khani, M., & Hosseini, S. E. (2013). The investigation of Lavidge and Steiner model's capability in measuring the advertising effectiveness of fire insurance. *European Online Journal of Natural and Social Sciences,2(3)*, 1559-1567.

Samsonova, E. V., Bäck, T., Kok, J. N., & IJzerman, A. P. (2005). Reliable hierarchical clustering with the self-organizing map. *International Symposium on Intelligent Data Analysis , 3646*, 385-396.

Samsonova, E. V., Kok, J. N., & IJzermana, A. P. (2006). TreeSOM: Cluster analysis in the self-organizing map. *Neural Networks, 19(6)*, 935-949.

Schiaffino, S., & Amandi, A. (2009). Intelligent User Profiling . *Artificial Intelligence An International Perspective*, 193-216.

Singh, A., Thakur, N., & Sharma, A. (2016). A Review of Supervised Machine Learning Algorithms. *EURASIP Journal on Advances in Signal Processing*, 67. doi:https://doi.org/10.1186/s13634-016-0355-x

Srimani, P., & Srinivas, A. (2011). Behavioral Targeting- Consumer Tracking. *AIP Conference Proceedings* , 56-60. doi: 10.1063/1.3669931

Steindel , T. A. (2010). A Path toward User Control of Online Profiling. *Mich. Telecomm. & Tech. L. Rev, 17(2)*, 459-489. Retrieved from https://repository.law.umich.edu/cgi/viewcontent.cgi?article=1036&amp=&context=mttlr&amp=&sei-redir=1&referer=https%253A%252F%252Fscholar.google.nl%252Fscholar%253Fstart%253D10%2526q%253Donline%252Bprofiling%2526hl%253Den%2526as_sdt%253D0%252C5#search=%22

Trusov, M., & Ma , L. (2016). Crumbs of the Cookie: User Profiling in Customer-Base Analysis and Behavioral Targeting. *Marketing Science, 35(3)*, 405–426. doi:http://dx.doi.org/10.1287/mksc.2015.0956

Tsai, C.-Y., & Chiu, C.-C. (2004). A purchase-based market segmentation methodology. *Expert Systems with Applications*, 265-276. doi:https://doi.org/10.1016/j.eswa.2004.02.005

Tu, S., & Lu, C. (2010). Topic-Based User Segmentation for Online Advertising with Latent Dirichlet Allocation. *International Conference on Advanced Data Mining and Applications*, 259-269. doi:https://doi.org/10.1007/978-3-642-17313-4_26

Univeristy of Twente. (2017, Jul. 28). *Fact&Figures Education* . Retrieved from University of Twente.: https://www.utwente.nl/en/facts-and-figures/education/

University of Twente. (2017, Jul. 14). *Electrical Engineering, Mathematics and Computer Science (EEMCS)*. Retrieved from University of Twente: https://www.utwente.nl/en/eemcs/

University of Twente. (2018, June 25). *PRIVACY: SECURE HANDLING OF PERSONAL DATA*. Retrieved from UNIVERSITY OF TWENTES.: https://www.utwente.nl/en/news/!/2018/6/204815/privacy-secure-handling-of-personal-data

Wiedmann , K.-P., Buxel , H., & Walsh, G. (2001). Customer profiling in e-commerce: Methodological aspects and challenges. *Journal of Database Marketing, 9(2)*, 170-184.

Retrieved from https://link.springer.com/content/pdf/10.1057%2Fpalgrave.jdm.3240073.pdf

Wijaya, B. S. (2011). The development of Hierachy of Effects Model in Advertising. *International research journal of Business Studies, 5(1)*, 73-85.

Wu, W.-y., & Ke, C.-c. (2015). An online shopping behavior model integrating personality traits, perceived risk, and technology acceptance. *Social Behavior and Personality, 43(1)*, 85-98. doi:10.2224/sbp.2015.43.1.85

Wu, X., Yan, J., Liu, N., Yan, S., Chen, Y., & Chen, Z. (2009). Probabilistic Latent Semantic User Segmentation for Behavioral Targeted Advertising. *Third International Workshop on Data Mining and Audience Intelligence for Advertising*, 10-17. doi:10.1145/1592748.1592751

Xie, J., Jiang, S., Xie, W., & Gao, X. (2011). An efficient global K-means clustering algorithm. *Journal of Computers, 6(2)*, 271-279.

Xing, E., Ho, Q., Xie, P., & Wei, D. (2016). Strategies and Principles of Distributed Machine Learning on Big Data. *Engineering, 2*, 179-195. doi:https://doi.org/10.1016/J.ENG.2016.02.008

Xiu, Y., Lan, M., Wu, Y., & Lang, J. (2017). Exploring Semantic Content to User Profiling for User Cluster-based Collaborative Point-of-Interest Recommender System. *International Conference on Asian Language Processing (IALP)* , 268-271 . doi:10.1109/IALP.2017.8300595

Yan, J., Liu, N., Wang, G., Zhang, W., Jiang, Y., & Chen, Z. (2009). How much can Behavioral Targeting Help Online Advertising? *WWW'09 - Proceedings of the 18th International World Wide Web Conference*, 261-270. doi:10.1145/1526709.1526745

Yang, S., Korayem , M., AlJadda , K., Grainger, T., & Natarajan , S. (2017). Combining content-based and collaborative filtering for job recommendation system: A cost-sensitive Statistical Relational Learning approach. *Knowledge-Based system, 136*, 37-45. doi:http://dx.doi.org/10.1016/j.knosys.2017.08.017

Yao, Z., Eklund, T., & Back, B. (2010). Using SOM-Ward clustering and predictive analytics for conducting customer segment. *International Conference on Data Mining Workshops*, 639-646.

Zahra, S., Ghazanfar, M. A., Khalid, A., & Azam, M. A. (2015). Novel centroid selection approaches for KMeans-clustering based recommender systems. *Information Sciences,320*, 156-189. doi:https://doi.org/10.1016/j.ins.2015.03.062

9. Appendix

Appendix 1 Machine Learning Algorithms (Brownlee, 2016)

Appendix 2 Statistics of each k-means cluster for non-admitted students

$`1`

	Cla/Mod	Mod/Cla	Global	p.value	v.test
Openday.request=NO Openday	51.519757	98.833819	76.869159	3.181722e-45	14.112484
PDF.download=NO PDF download	52.180685	97.667638	75.000000	5.747502e-44	13.906955
Studielink.application=NO Studielink	50.301205	97.376093	77.570093	3.737190e-36	12.554886
E-check=NO Echeck	50.623053	94.752187	75.000000	7.772306e-32	11.741897
Scholarship.finder=NO Scholarship	47.391952	92.711370	78.387850	1.729703e-18	8.773647
opleiding=NO Program check	42.555831	100.000000	94.158879	2.779102e-12	6.988465
Questionsviawebforms=NO	42.241379	100.000000	94.859813	7.587650e-11	6.508556
algemeen=NO Master check	40.639810	100.000000	98.598131	2.037338e-03	3.084735
algemeen=Master check	0.000000	0.000000	1.401869	2.037338e-03	-3.084735
Questionsviawebforms=Questions	0.000000	0.000000	5.140187	7.587650e-11	-6.508556
opleiding=Program check	0.000000	0.000000	5.841121	2.779102e-12	-6.988465
Scholarship.finder=Scholarship	13.513514	7.288630	21.612150	1.729703e-18	-8.773647
E-check=Echeck	8.411215	5.247813	25.000000	7.772306e-32	-11.741897
Studielink.application=Studielink	4.712042	2.623907	22.313084	7.162114e-36	-12.503294
PDF.download=PDF download	3.738318	2.332362	25.000000	5.747502e-44	-13.906955
Openday.request=Openday	2.020202	1.166181	23.130841	3.181722e-45	-14.112484

$`2`

	Cla/Mod	Mod/Cla	Global	p.value	v.test
Openday.request=Openday	76.262626	50.166113	23.130841	2.982127e-42	13.621540
PDF.download=PDF download	69.158879	49.169435	25.000000	2.910215e-32	11.824677
educationbrochure=NO Educational Brochure	43.655589	96.013289	77.336449	4.259150e-26	10.566541
Studielink.application=Studielink	63.350785	40.199336	22.313084	1.344197e-19	9.056730
Program =NO Program check	37.344913	100.000000	94.158879	1.723409e-10	6.384160
Questionsviawebform=NO	37.068966	100.000000	94.859813	2.805128e-09	5.942614
Master =NO Master check	35.663507	100.000000	98.598131	5.290060e-03	2.788819
Scholarship.finder=NO Scholarship	37.406855	83.388704	78.387850	8.229722e-03	2.642497
Scholarship.finder=Scholarship	27.027027	16.611296	21.612150	8.229722e-03	2.642497
Master =Master check	0.000000	0.000000	1.401869	5.290060e-03	2.788819
Questionsviawebform=Quesitions	0.000000	0.000000	5.140187	2.805128e-09	5.942614
Program =Program check	0.000000	0.000000	5.841121	1.723409e-10	6.384160
Studielink.application=NO Studielink	27.108434	59.800664	77.570093	2.670214e-19	8.981532
educationbrochure=Educational Brochure	6.185567	3.986711	22.663551	4.259150e-26	10.566541
PDF.download=NO PDF download	23.831776	50.830565	75.000000	2.910215e-32	11.824677
Openday.request=NO Openday	22.796353	49.833887	76.869159	2.982127e-42	13.621540

```
$`3`
                                Cla/Mod Mod/Cla    Global      p.value   v.test
Questionsviawebform=Questions      100     100  5.140187 7.656903e-75 18.30423
Questionsviawebform=NO               0       0 94.859813 7.656903e-75 18.30423
```

```
$`4`
                                          Cla/Mod     Mod/Cla    Global      p.value    v.test
E-check=Echeck                          44.392523  81.1965812 25.000000 1.899897e-44
13.985925
Scholarship.finder=Scholarship          45.945946  72.6495726 21.612150 1.160244e-38 13.004055
educationbrochure=Educational Brochure  42.783505  70.9401709 22.663551 3.294379e-34 12.195263
Questionsviawebform=NO                  14.408867 100.0000000 94.859813 1.296057e-03  3.216851
opleiding=NO Program check              14.392060  99.1452991 94.158879 5.322660e-03  2.786829
opleiding=Program check                  2.000000   0.8547009  5.841121 5.322660e-03  2.786829
Questionsviawebform=Question             0.000000   0.0000000  5.140187 1.296057e-03  3.216851
educationbrochure=NO Educational Brochure  5.135952  29.0598291 77.336449 3.294379e-34 12.195263
Scholarship.finder=NO Scholarship        4.769001  27.3504274 78.387850 1.160244e-38 13.004055
E-check=NO Echeck                        3.426791  18.8034188 75.000000 1.899897e-44 13.985925
```

```
$`5`
                                     Cla/Mod Mod/Cla    Global      p.value     v.test
opleiding=Program check            86.0000000      86  5.841121 1.332013e-57  15.997410
algemeen=Master check             100.0000000      24  1.401869 4.059808e-16   8.136765
Studielink.application=Studielink  10.4712042      40 22.313084 3.867526e-03   2.888769
E-check=Echeck                      9.3457944      40 25.000000 1.651155e-02   2.397408
E-check=NO Echeck                   4.6728972      60 75.000000 1.651155e-02  -2.397408
Studielink.application=NO Studielink   4.5180723      60 77.570093 4.147032e-03  -2.866756
algemeen=NO Master check            4.5023697      76 98.598131 4.059808e-16  -8.136765
opleiding=NO Program check          0.8684864      14 94.158879 1.332013e-57 -15.997410
```

```
$`6`
                                              Cla/Mod Mod/Cla    Global     p.value   v.test
Studielink.application=Studielink application     100     100 0.1168224 0.0011682243.246527
```

Appendix 3 Statistics of each k-means cluster for admitted student

$`1`

	Cla/Mod	Mod/Cla	Global	p.value	v.test
Program webpage= Program webpage	55.737705	100.000000	81.024668	6.211271e-31	11.564832
Openday=openday	57.462687	97.058824	76.280835	3.114845e-28	11.018418
scholarship=scholarship	57.329843	92.016807	72.485769	1.773017e-21	9.517555
PDF.download=NO PDF. download	51.293103	100.000000	88.045541	1.166382e-18	8.817889
Check.admission.requirement	58.461538	79.831933	61.669829	2.859297e-15	7.896898
Education.brochure= Education Brochure	51.441242	97.478992	85.578748	3.576187e-14	7.575547
masterwebpage= masterwebpage	50.843242	60.448792	80.578258	3.576187e-14	7.075585
Eligibility.check=NO Echeck	53.333333	87.394958	74.003795	8.287871e-11	6.495281
Questionsviawebform=NO	48.080808	100.000000	93.927894	1.967486e-09	6.000469
Questionsviawebform=Question	0.000000	0.000000	6.072106	1.967486e-09	-6.000469
Eligibility.check=Echeck	21.897810	12.605042	25.996205	8.287871e-11	-6.495281
PDF.download=PDF download	0.000000	0.000000	11.954459	1.166382e-18	-8.817889
masterwebpage= no masterwebpage	13.103448	7.983193	27.514231	1.773017e-21	-9.517555
Check.admission.requirement= no	5.600000	2.941176	23.719165	3.114845e-28	-11.018418
opleiding=NO opleiding	0.000000	0.000000	18.975332	6.211271e-31	-11.564832

$`2`

	Cla/Mod	Mod/Cla	Global	p.value	v.test
Program selection=program selection	77.000000	79.381443	18.975332	3.141228e-52	15.207757
Scholarship=Scholarship	34.297521	85.567010	45.920304	5.535170e-19	8.900988
Program webpage= Program webpage	48.684211	35.576923	14.421252	4.862148e-10	6.223490
PDF.download=NO PDF download	20.905172	100.000000	88.045541	1.086148e-06	4.875352
Questionsviawebform=NO	19.595960	100.000000	93.927894	1.192724e-03	3.240614
Education.brochure=NO Education Brochure	19.733925	91.752577	85.578748	4.888035e-02	1.969634
Education.brochure=Education brochure	10.526316	8.247423	14.421252	4.888035e-02	-1.969634
Questionsviawebform=Questions	0.000000	0.000000	6.072106	1.192724e-03	-3.240614
PDF.download=PDF download	0.000000	0.000000	11.954459	1.086148e-06	-4.875352

$`3`

	Cla/Mod	Mod/Cla	Global	p.value	v.test
scholarship=scholarship	38.613861	75.000000	38.330171	2.128319e-17	8.486566
Openday no openday	48.684211	35.576923	14.421252	4.862148e-10	6.223490
Check.admission.requirement	41.600000	53.608247	23.719165	8.546361e-13	7.152095
scholarship= no scholarship	8.000000	25.000000	61.669829	2.128319e-17	-8.486566
Check.admission.requirement=no	8.638743	31.730769	72.485769	7.433569e-23	-9.841844

$`4`

	Cla/Mod	Mod/Cla	Global	p.value	v.test
Program webpage= Program webpage	48.965517	68.269231	27.514231	7.433569e-23	9.841844
Openday=openday	38.613861	75.000000	38.330171	2.128319e-17	8.486566
Openday= no openday	0.000000	0.00000	6.072106	2.443298e-02	-2.250252
Program webpage= no Program webpage	8.872902	66.07143	79.127135	1.620855e-02	-2.404185

$`5`

	Cla/Mod	Mod/Cla	Global	p.value	v.test
Eligibility.check=Eligibility.check	100.00000	100.000	6.072106	5.487520e-52	15.171187
Program webpage= Program webpage	55.43857	98.05824	76.280835	3.114845e-28	11.018418
Openday=openday	50.737705	85.41572	73..4668	3.011271e-31	10.564832
scholarship=scholarship	57.329843	92.016807	72.485769	1.773017e-21	9.517555
Education.brochure=NO Education Brochure	15.328467	37.50000	25.996205	4.544162e-02	2.000543
scholarship= no scholarship	0.000000	0.000000	11.954459	1.086148e-06	-4.875352
Openday=no openday	11.194030	46.391753	76.280835	8.546361e-13	-7.152095
Program webpage= no Program webpage	4.912281	14.432990	54.079696	5.535170e-19	-8.900988
Eligibility.check= no Eligibility.check	4.683841	20.618557	81.024668	3.141228e-52	-15.207757

Appendix 4 User profile tables

Table 15 Distribution of Interest country ISO code

	Cluster number						
Interest country ISO code	**1**	**2**	**3**	**4**	**5**	Total	%
BQ				1		**1**	0,19%
BR		1				**1**	0,19%
CN		1				**1**	0,19%
CO				1		**1**	0,19%
CZ			1		1	**1**	0,19%
DE	2	23	8	15	4	**52**	9,85%
DK	1					**1**	0,19%
EC				1		**1**	0,19%
EE	1					**1**	0,19%
EG		1			1	**2**	0,38%
ES	1	1				**2**	0,38%
FR				1		**1**	0,19%
GB		2		1		**3**	0,57%
GH		1			1	**2**	0,38%
GR	1	1	2	2		**6**	1,14%
IN	1	2	5	6	4	**18**	3,41%
IR				1		**1**	0,19%
IT		3				**3**	0,57%
KR				1		**1**	0,19%
LB					1	**1**	0,19%
LV		1				**1**	0,19%
MX			1	1		**2**	0,38%
MY		1				**1**	0,19%
NG			1			**1**	0,19%
NL	62	75	57	134	60	**388**	73,48%
PK					2	**2**	0,38%
PL		2				**2**	0,38%
PS				1		**1**	0,19%
PT	1					**1**	0,19%
RO	1		1			**2**	0,38%
RU			1			**1**	0,19%
SE				1		**1**	0,19%
SI					1	**1**	0,19%
SR	1			1		**2**	0,38%
SY		1	1			**2**	0,38%
TR				2		**2**	0,38%
UA				1		**1**	0,19%
US	2	1	1	1		**5**	0,95%
XK	1			1		**2**	0,38%
N.A.	1	3	3	3		**10**	1,89%

Total	**76**	**120**	**82**	**176**	**75**	**528**	**100%**

Table 16 Distribution of source/medium

	Cluster number						
Source/medium	1	2	3	4	5	**Total**	%
(direct)/(non)	2	8	10	15	4	39	7,39%
SEintake2017leads/email		1				1	0,19%
app.studielink.nl/referral			1			1	0,19%
bing/cpc				2		2	0,38%
bing/organic		1		1		2	0,38%
com.google.android.gm/referral			1			1	0,19%
de.search.yahoo.com		1		1		2	0,38%
deref-web-02.de/refer				2		2	0,38%
doorstroommatrix.nl/refer	1			1		2	0,38%
ecosia.org/referral				1		1	0,19%
edmundo.ro/refrral	1					1	0,19%
ethicsandtechnology.eu/refer		1				1	0,19%
facebook.com/cpc	1	1		1		3	0,57%
facebook.com/referral		1				1	0,19%
follow-up-od-mrt-17/email			1			1	0,19%
follow-up-od-nov16/email			1			1	0,19%
google/cpc	6	10	3	10	5	34	6,44%
google/organic	57	77	54	113	53	354	67,05%
info.studielink.nl/referral		1				1	0,19%
keuzegids.org/referral		1				1	0,19%
kick-in.nl/referral			1			1	0,19%
l.facebook.com/referral				1		1	0,19%
linkedin.com/referral	1					1	0,19%
mail.google.com/referral		1		1	1	3	0,57%
mastersportal.eu/referral		1		1		2	0,38%
nieuwsbrief-vooraannmelders/emails	1			1	1	3	0,57%
nl/search.yahoo.com/referral				1		1	0,19%
outlook/live.com/referral	3	3	1	5	8	20	3,79%
qa-extra-uitnodiging/email			1	1	1	3	0,57%
quicklink/print	2	1		3	1	7	1,33%
rooster.utwente.nl/referral				1		1	0,19%
sigon.utwente.nl/referral		1		1		2	0,38%
ssl1.peoplexs.com/referral			1			1	0,19%
staticxx.facebook.com		1	1	2	1	5	0,95%
studiekeuze123.nl/referral				1		1	0,19%
studyinholland.nl/referral		1				1	0,19%
topuniversities.com/referral				2		2	0,38%
twente.com/referral				1		1	0,19%
uitnodigingsmail-bsc/email		1				1	0,19%

uitnodigingsmail-msc/email	1	1		1		3	0,57%
Universitaire.bachelor.nl/referral		1				1	0,19%
utwente.nl/referral			1			1	0,19%
vest.nl/referral		1				1	0,19%
N.A.		4	5	5		14	2,65%
Total	76	120	82	175	75	528	100,00%

Table17 targeted student distribution

Faculty	Source/Medium	ISO code	Education Level	Cluster number 1	2	3	4	5	Total	%
BMS	(direct)/(non)	DE	MSc		1		1	1	3	0,57%
		NL	BSc		1		2		3	0,57%
			MSc		3	2	2	1	8	1,52%
	app.studielink.nl/refer	DE	BSc			1			1	0,19%
		NL	BSc				1		1	0,19%
	bing/cpc	NL	BSc				1		1	0,19%
	bing/organic	DE	BSc		1				1	0,19%
		NL	BSc				1		1	0,19%
	de.search.yahoo.com	NL	BSc		1				1	0,19%
	deref-web-02.de/refer	US	BSc				1		1	0,19%
	doorstroommatrix.nl/refer	NL	BSc				1		1	0,19%
	ethicsandtechnology.eu/refer	IT	MSc		1				1	0,19%
	facebook.com/referral	NL	MSc		1				1	0,19%
	facebook.com/cpc	NL	MSc					1	1	0,19%
	follow-up-od-mrt-17/email	DE	MSc			1			1	0,19%
	follow-up-od-nov16/email	NL	BSc			1			1	0,19%
	google/cpc	DE	BSc		2				2	0,38%
		IT	MSc		1				1	0,19%
		NL	BSc	1	1		1		3	0,57%
			MSc	2	1				3	0,57%
	google/organic	CZ	BSc			1			1	0,19%
			MSc					1	1	0,19%
		DE	BSc		5	3	11	1	20	3,79%
			MSc	1	7	2	1	1	12	2,27%
		ES	MSc		1				1	0,19%
		FR	BSc				1		1	0,19%
		GB	BSc		1				1	0,19%
		NL	BSc	14	11	8	22	11	66	12,50%
			MSc	5	4	8	10	13	40	7,58%

		RU	MSc			1			1	0,19%
		SE	BSc				1		1	0,19%
	mail.google.com/refer	DE	MSc		1				1	0,19%
	nieuwsbrief-vooraannmelders/emails	DE	BSc				1		1	0,19%
	outlook.live.com/refer	EG	BSc					1	1	0,19%
		NL	BSc	1			2	3	6	1,14%
			MSc	1		1			2	0,38%
	sigon.utwente.nl/referral	DE	MSc		1				1	0,19%
		NL	BSc				1		1	0,19%
	staticxx.facebook.com/referral	CN	MSc		1				1	0,19%
		DE	BSc				1		1	0,19%
	uitnodigingsmail-msc	ES	MSc	1					1	0,19%
		NL	MSc				1		1	0,19%
	Universitaire.bachelor	NL	BSc		1				1	0,19%
	utwente.nl/referral	NL	MSc			1			1	0,19%
	veste.nl/referral	NL	MSc		1				1	0,19%
	N.A.	NL	MSc		1				1	0,19%
		N.A.	BSc		1				1	0,19%
			MSc			1	2		3	0,57%
EWI	(direct)/(non)	NL	BSc				1		1	0,19%
			MSc		1	2	3		6	1,14%
	edmundo.ro/refrral	RO	MSc	1					1	0,19%
	google/cpc	NL	BSc	1		1	2	1	5	0,95%
			MSc		2				2	0,38%
	google/organic	BR	BSc		1				1	0,19%
		DE	BSc	1	2		1		4	0,76%
		DK	BSc	1					1	0,19%
		EE	MSc	1					1	0,19%
		GR	BSc	1					1	0,19%
		IN	BSc				1		1	0,19%
			MSc		1			3	4	0,76%
		MY	MSc	1					1	0,19%
		NL	BSc	7	7	3	10	2	29	5,49%
			MSc	4	4	2	4	1	15	2,84%
		PL	BSc		1				1	0,19%
		RO	BSc			1			1	0,19%
		SI	BSc					1	1	0,19%
		TR	MSc				1		1	0,19%
		US	BSc			1			1	0,19%
		N.A.	MSc	1					1	0,19%
	keuzegids.org/referral	NL	MSc		1				1	0,19%
	l.facebook.com/referral	XK	BSc				1		1	0,19%

	linkedin.com/referral	NL	BSc	1					1	0,19%
	mail.google.com/referral	NL	MSc				1		1	0,19%
	nieuwsbrief-vooraannmelders/emails	DE	BSc					1	1	0,19%
	outlook.live.com/referral	NL	BSc				1		1	0,19%
	studiekeuze123.nl/referral	BQ	BSc				1		1	0,19%
	topuniversities.com/referral	TR	MSc				1		1	0,19%
		UA	BSc				1		1	0,19%
	twente.com/referral	DE	BSc				1		1	0,19%
	uitnodigingsmail-bsc/email	PL	BSc		1				1	0,19%
ET	(direct)/(non)	GR	MSc			1			1	0,19%
		IN	MSc	1		1			2	0,38%
	SEintake2017leads/email	NL	MSc		1				1	0,19%
	bing/pcp	KR	BSc				1		1	0,19%
	doorstroommatrix.nl/referral	NL	MSc					1	1	0,19%
	ecosia.org/referral	NL	MSc				1		1	0,19%
	facebook/cpc	DE	MSc		1				1	0,19%
		NL	MSc	1					1	0,19%
	goodle/cpc	EG	BSc		1				1	0,19%
		NL	MSc	1	1			1	3	0,57%
	google/organic	DE	BSc		1		1		2	0,38%
		GB	MSc					1	1	0,19%
		GR	MSc				1		1	0,19%
		IN	MSc		1	3	2	1	7	1,33%
		IT	MSc		1				1	0,19%
		MX	MSc			1	1		2	0,38%
		NL	BSc	10	5	6	12	6	39	7,39%
			MSc	3	7	2	2	2	16	3,03%
		US	BSc		1				1	0,19%
			MSc	1					1	0,19%
	info.studielink.nl/referral	LV	MSc		1				1	0,19%
	mastersportal.eu/referral	GB	MSc		1				1	0,19%
		PS	MSc				1		1	0,19%
	outlook.live.com/referral	NL	MSc				2		2	0,38%
		SR	MSc	1					1	0,19%
	qa-extra-uitnodiging/email	NL	BSc			1	1		2	0,38%
	quicklink/print	GR	MSc		1				1	0,19%
		NL	BSc				1	1	2	0,38%
	rooster.utwente.nl/referral	NL	MSc				1		1	0,19%
	staticxx.facebook.com/referral	IN	MSc				1		1	0,19%
		PK	MSc					1	1	0,19%

	studyinholland.nl/referral	SY	MSc		1				1	0,19%
	N.A.	GB	MSc				1		1	0,19%
		NL	BSc			2			2	0,38%
		N.A.	BSc		1		1		2	0,38%
TNW	(direct)/(non)	CO	MSc				1		1	0,19%
		NL	BSc			2	1	1	4	0,76%
			MSc		1	2	5	1	9	1,70%
		US	MSc	1					1	0,19%
	com.google.android.gm/referral	NL	BSc			1			1	0,19%
	foostroommatrix.nl/referral	NL	BSc	1			1		2	0,38%
	facebook/cpc	NL	BSc				1		1	0,19%
	google/cpc	IN	MSc				1		1	0,19%
		NL	BSc	1	1	1	6	3	12	2,27%
	google/organic	DE	BSc		2		1		3	0,57%
		EC	BSc				1		1	0,19%
		GR	MSc				1		1	0,19%
		IN	MSc			1	1		2	0,38%
		IR	MSc				1		1	0,19%
		NL	BSc	3	7	7	13	3	33	6,25%
			MSc	1	7	4	7	2	21	3,98%
		SR	BSc				1		1	0,19%
	kick-in.nl/referral	NL	BSc			1			1	0,19%
	mail.google.com/referral	NL	MSc					1	1	0,19%
	nieuwsbrief-vooraannmelders/emails	NL	MSc	1					1	0,19%
	nl/search.yahoo.com/referral	NL	MSc				1		1	0,19%
	outlook/live.com/referral	NL	BSc		3			2	5	0,95%
	qa-extra-uitnodiging/email	NL	BSc					1	1	0,19%
	quicklink/print	NL	BSc	1			2		3	0,57%
		PT	MSc	1					1	0,19%
	ssl1.peoplexs.com/referral	NL	MSc			1			1	0,19%
	staticxx.facebook.com	GR	MSc			1			1	0,19%
	N.A.	NL	BSc			1			1	0,19%
		N.A.	BSc		1	2			3	0,57%
N.A.	(direct)/(non)	NG	MSc		1				1	0,19%
	google/organic	GH	MSc		1				1	0,19%
		LB	MSc					1	1	0,19%
		NL	MSc	2			4	2	8	1,52%
		PK	MSc					1	1	0,19%
	quicklink/print	NL	MSc				1		1	0,19%
Total				76	120	82	175	75	528	100,00%