**NATIONAL UNIVERSITY HO CHI MINH CITY**

**UNIVERSITY OF INFORMATION TECHNOLOGY  
ADVANCED PROGRAM IN INFORMATION SYSTEMS**

**DAO MINH HIEU – 15520219**

**HOANG TRONG VIET – 15521009**

ARTIFICIAL INTELLIGENT

IN MARKET SEGMENTATION

AND

APPLICATION

**BACHELOR OF ENGINEERING IN INFORMATION SYSTEMS**

**THESIS ADVISOR**

**Professor: NGUYEN DINH THUAN**

**ASSESSMENT COMMITTEE**

|  |  |  |
| --- | --- | --- |
| The Assessment Committee is established under the Decision ................ | | , date |
| ................... by Rector of the University of Information Technology. | |  |
| 1 .................................................. | - Chairman. |  |
| 2 .................................................. | - Secretary. |  |
| 3 .................................................. | - Member |  |
| 4 .................................................. | - Member. |  |

**ACKNOWLEDGEMENTS**

1

**TABLE OF CONTENT**

|  |  |  |  |
| --- | --- | --- | --- |
| ACKNOWLEDGEMENTS………....................................................................................... | | | x |
| TABLE OF CONTENT ........................................................................................................ | | | x |
| LIST OF FIGURES ............................................................................................................... | | | x |
| LIST OF TABLES ................................................................................................................ | | | x |
| ABSTRACT ........................................................................................................................... | | | x |
| Chapter 1 | | INTRODUCTION .......................................................................................... | x |
| 1.1 | Overview ............................................................................................................. | | x |
| 1.2 | Project Description .............................................................................................. | | x |
| 1.3 | Scope & Limitation ............................................................................................. | | x |
| 1.4 | Thesis structure ................................................................................................... | | x |
| Chapter 2 FUNDAMENTAL THEORIES AND TECHNOLOGIES USED ............... | | | x |
| 2.1 | Marketing & Digital Marketing .......................................................................... | | x |
| 2.2 | Marketing Tools .................................................................................................. | | x |
| 2.3 | AI –ML in Customer Segmentation .................................................................... | | x |
| 2.4 | RFM Analysis ..................................................................................................... | | x |
| 2.5 | Clustering Algorithm .......................................................................................... | | x |
| 2.6 | ASP.NET Core..................................................................................................... | | x |
| 2.7 | Front-end Technologies ...................................................................................... | | x |
| 2.8 | Back-end Technologies ....................................................................................... | | x |
| Chapter 3 APPLICATION ANALYSIS AND DESIGN ................................................ | | | x |
| 3.1 | Requirements ...................................................................................................... | | x |
| 3.2 | Architecture ......................................................................................................... | | x |
| 3.3 | Use-case Diagram ............................................................................................... | | x |
| 3.4 | Class Diagram ..................................................................................................... | | x |
| 3.5 | Activity Diagram ................................................................................................ | | x |
| 3.6 | UI Design Using WireframePro .......................................................................... | | x |
| Chapter 4 MARKETING APPLICATION IMPLEMENTATION ……………………… | | | x |
| Chapter 5 EVALUATION ……………………………................................................ | | | x |
| Chapter 6 CONCLUSION AND FUTURE WORKS ….............................................. | | | x |
| 6.1 | Conclusion .......................................................................................................... | | x |
| 6.2 | Future works ....................................................................................................... | | x |
| REFERENCES ……………………………………….….............................................. | | | x |

**LIST OF FIGURES**

*Figure 2.1. Search result on the Internet.* *x*

*Figure 2.2. Most popular activities on smartphones.* *x*

*Figure 2.3. Rating of effectiveness of digital marketing channels.* *x*

*Figure 2.4. AI & ML in marketing.* *x*

*Figure 2.5. RFM Metrics.* *x*

*Figure 2.6. K-means algorithm.* *x*

*Figure 2.7. K-means visualization.* *x*

*Figure 2.8. ASP.NET Core advantages.* *x*

*Figure 2.9. Architecture.* *x*

*Figure 3.1. Use case diagram.* *x*

*Figure 3.2. Class diagram.* *x*

*Figure 3.3. Activity diagram.* *x*

*Figure 3.4. Login screen.* *x*

*Figure 3.5. Dashboard screen.* *x*

*Figure 3.6. Mailing list screen.* *x*

*Figure 3.7. My campaign screen.* *x*

*Figure 3.8. Add template screen.* *x*

*Figure 3.9. Template library screen.* *x*

*Figure 4.1. Create campaign step 1.* *x*

*Figure 4.2. Create campaign step 2.* *x*

*Figure 4.3. Create campaign step 3.* *x*

*Figure 4.4. Create campaign step 4.* *x*

*Figure 4.5. Create campaign step 5.* *x*

*Figure 4.6. Create campaign step 6.* *x*

*Figure 4.7. Error page.* *x*

**LIST OF TABLES**

*Table 1. Email marketing tools comparison*  *x*

*Table 2. Recommended segments for RFM* *x*

*Table 3. Comparison between algorithm* *x*

**ABSTRACT**

**Chapter 1** **INTRODUCTION**

**1.1** **Overview**

Data and technology have already transformed the way business engage their customers in the 21st century, Digital Marketing speeds up the race between Marketers to find their potential customers.

There are four main strategies are widely applied all over the world such as Affiliate Marketing, E-mail Marketing, Keyword Advertising and Online Advertising [1]. Take a closer look:

* Affiliate Marketing, it’s a way to promote products or services for merchants, frees up more money for advertising and enables retailers to save money on advertisements for sales than they would do for traditional online advertisements.
* Email Marketing means to contact customers directly through e-mails, it’s a very popular strategy because of saving cost. Moreover, e-mails can also show one or more pictures about the promoted products but the companies have to face the problem of spam filters because these filters cannot distinguish between requested and unrequested e-mails.
* Keyword Advertising aims to put paid banners on results pages of search engines. It became more meaningful in the world of online marketing for companies because 45.9% of all interviewees did find a new online shop while using a searching engine.
* Online Advertising is one of the oldest strategies but it also the most used advertisement in the online world because of decreasing prices and increasing quality of videos, pictures, etc.

Digital Marketing can develop long-term and pleasant relationships with customers if they can detect and predict changes in customer behavior. In the dynamic retail market, understanding changes in customer behavior can help marketers to establish effective promotion campaigns [2]

Digital marketing engagement can be categorized according to perceived benefits and Digital Marketing usage. A business is nothing without its people. Marketers are forced to adapt their online strategy to increase customer attraction by focusing on relationship-based interactions with their clients [3].

We believe that Digital Marketing is an attractive proposition for many more consumer product or service categories than is typically assumed. Digital Marketing should fit with interactive media and its attractiveness for marketing-oriented relationship building, suggests a breadth of opportunity for Digital Marketing across many more categories than common wisdom would suggest [4].

Therefore, we decide to build this tool to take action on customer insight to help Marketers to offer better, more relevant and engaging content to customers. We optimize E-mail Marketing and Online Marketing by studying about customer behaviors to support decision making for Marketers.

**1.2** **Project Description**

A web application based on Market Segmentation to analyze customer behavior to support for Marketers easier to suggest proper products for their customers. The primary purpose of the application is to divide whole customer data into different segments. Each segment is distinct from the others and is comprised of client with similar needs, preferences and learning ability. We implement K-means method to create a customer insight and furthermore, we also use R to visualize and generate statistic from the dataset which was collected and combined from data of Vietnam and UCI website.

* 1. **Scope and Limitation**
* Data processing online.
* RFM file processing.
* Using K-means rules to mine quantitative data frequently present in database.
* Audience targeting.
* Web development.
* Limitation depends on customer data.
  1. **Thesis structure**

This thesis is divided into six chapters, as follows:

* **Chapter 1:** Introduction
* **Chapter 2:** Fundamental theories and technologies used
* **Chapter 3:** Application Analysis and Design
* **Chapter 4:** Implementation
* **Chapter 5:** Evaluation
* **Chapter 6:** Conclusion and Future Works.

**Chapter 2 FUNDAMENTAL THEORIES AND TECHNOLOGIES USED**

**2.1** **Marketing & Digital Marketing**

2.1.1 About Marketing

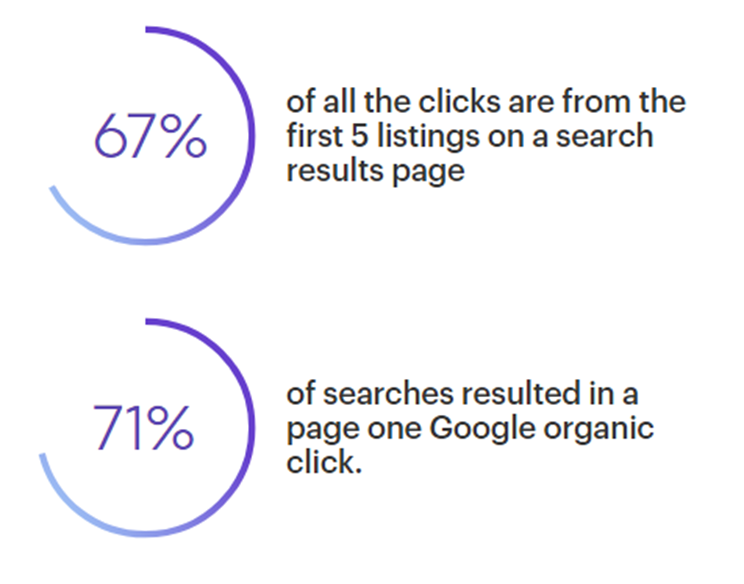
Marketing is the activity, set of institutions, and processes for creating, communicating, delivering, and exchanging offerings that have value for customers, clients, partners, and society at large1.

The science and art of exploring, creating, and delivering value to satisfy the needs of a target market at a profit. Marketing identifies unfulfilled needs and desires. It defines, measures and quantifies the size of the identified market and the profit potential. It pinpoints which segments the company is capable of serving best and it designs and promotes the appropriate products and services2.

2.1.2 About Digital Marketing

The use of Internet and other digital media and technology to support “modern marketing” has given rise to a bewildering range of labels and jargon created by both academics and professionals. It has been called Digital Marketing.

Surveys on the Internet show that:

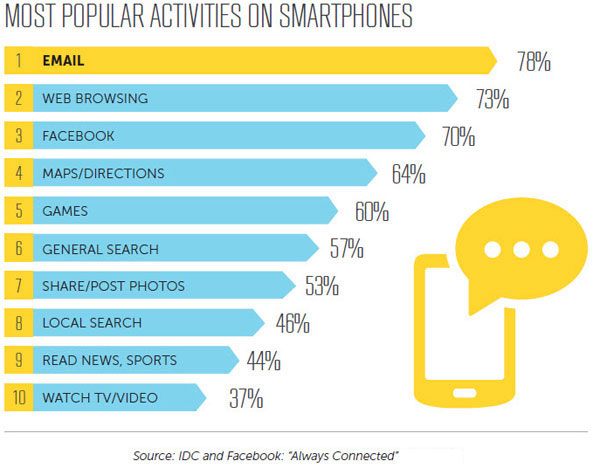


*Figure 2.1. Search result on the Internet.*

2.1.3 About Email Marketing

There are many types of digital marketing, but in our project we just consider the E-mail marketing.

* E-mail marketing is the use of email to develop relationships with potential customers and/or clients. Email marketing is one segment of internet marketing, which encompasses online marketing via websites, social media, blogs, etc. It is essentially the same as direct mail except that instead of sending mail through the postal service, messages are sent electronically via email.



*Figure 2.2. Most popular activities on smartphones.*

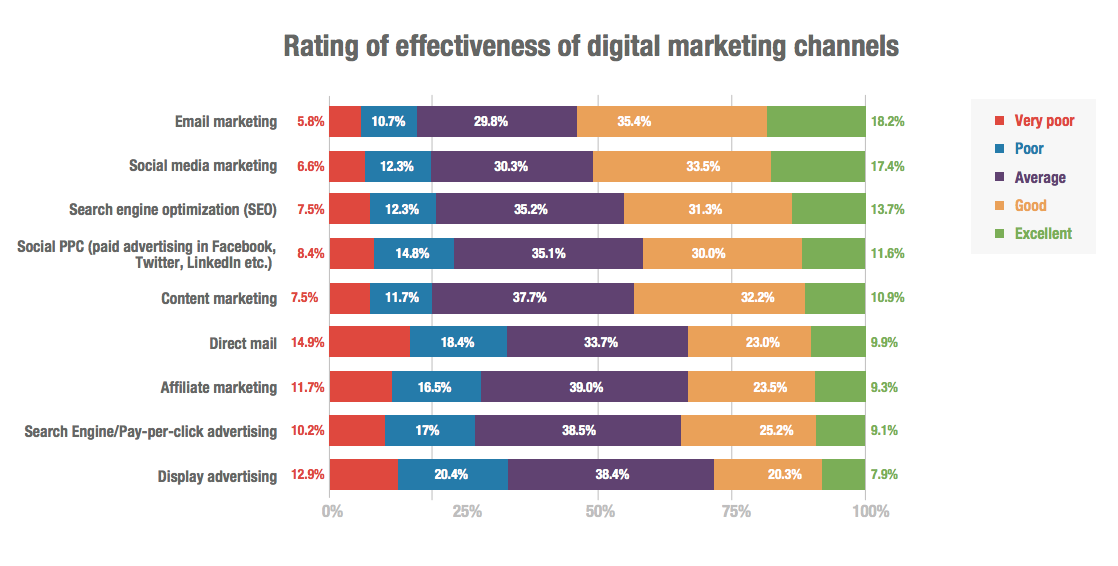
The two big advantages of email marketing are price and ease. Emailing is an inexpensive way to advertise your company and its products and/or services compared to many other types of marketing. It's also extremely easy to set up and track an email marketing campaign, making it a very accessible type of marketing for small businesses.

Newsletters can be sent to the email list you've built from the people who provided the necessary information on your website, for instance, providing these potential customers with news updates about your company, upcoming events and/or special offers – and, of course, reminding them that your business exists and that maybe it's time for another visit.

In an age of increasing usage of social media for advertising, email marketing still rules the roost, according to a study conducted by HostPapa:

* 94% of Internet users use email, while only 61% use social media.
* 75% of adult online users say that email marketing is their preferred marketing method.
* The "opt-in" feature of email allows marketing by consent.
* E-mail marketing allows targeting by demographics (age, income, etc.).
* E-mail messages have a much wider range of formatting possibilities than social media messages.
* E-mail offers more reporting and analytics capability, such as click-through rates, open rates, bounce rates, and conversions.

The huge advantage of email over social media is that prospects and customers are more likely to see an email than social media.



*Figure 2.3. Rating of effectiveness of digital marketing channels.*

**2.2** **Marketing Tools**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | **Constant Contact** | **Drip** | **ConvertKit** | **SendinBlue** |
| **Ease of use** | Recommended for smart businesses who’re just starting out with email marketing | Recommended for eCommerce websites and savvy internet marketers. | Recommended for publishers | Recommended for beginner marketers |
| **Automation** | - Customized content and promotion ideas based on selected industry.  - Offers unmatched support with live chat, emails, and discussion forums. They also have a vast library of resources. | - Convert visitors into subscribers and customers with intelligent marketing automation  - Deeper interaction with customers and send emails at the right time by understanding their purchase intent and behavior. | - Makes email automation easy to understand for publishers.  - Create attractive forms and send custom-tailored emails to individual subscribers. | - Send transactional messages with reliable delivery based on customer interaction. |
| **Industry** | Small business | Big Organization | Only for bloggers and publishers | Small Business only. |
| **Price** | $95/month – For 5001 – 10,000 subscribers. | Price varies and includes all premium features,  dedicated account manager, priority support, and more. | The cost varies for this plan and is for people who have more than 5000 active subscribers. | $79/month – For 3000 – 5000 subscribers and all features. |

*Table 1. Email marketing tools comparison.*

**2.3** **AI & ML in Customer Segmentation**

Customer segmentation is the practice of dividing a customer base into groups of individuals that are similar in specific ways relevant to marketing, such as age, gender, interests and spending habits.

2.3.1 Artificial Intelligence in Marketing

For marketers to reach their customers with the level of personalization that many have come to expect, they need to target increasingly granular segments.

AI can be used to achieve this. Drawing on the data that marketers already have about their customers, machine learning algorithms can be trained against a “gold standard” training set to identify important variables and common properties, and even pick out incorrectly identified contacts.

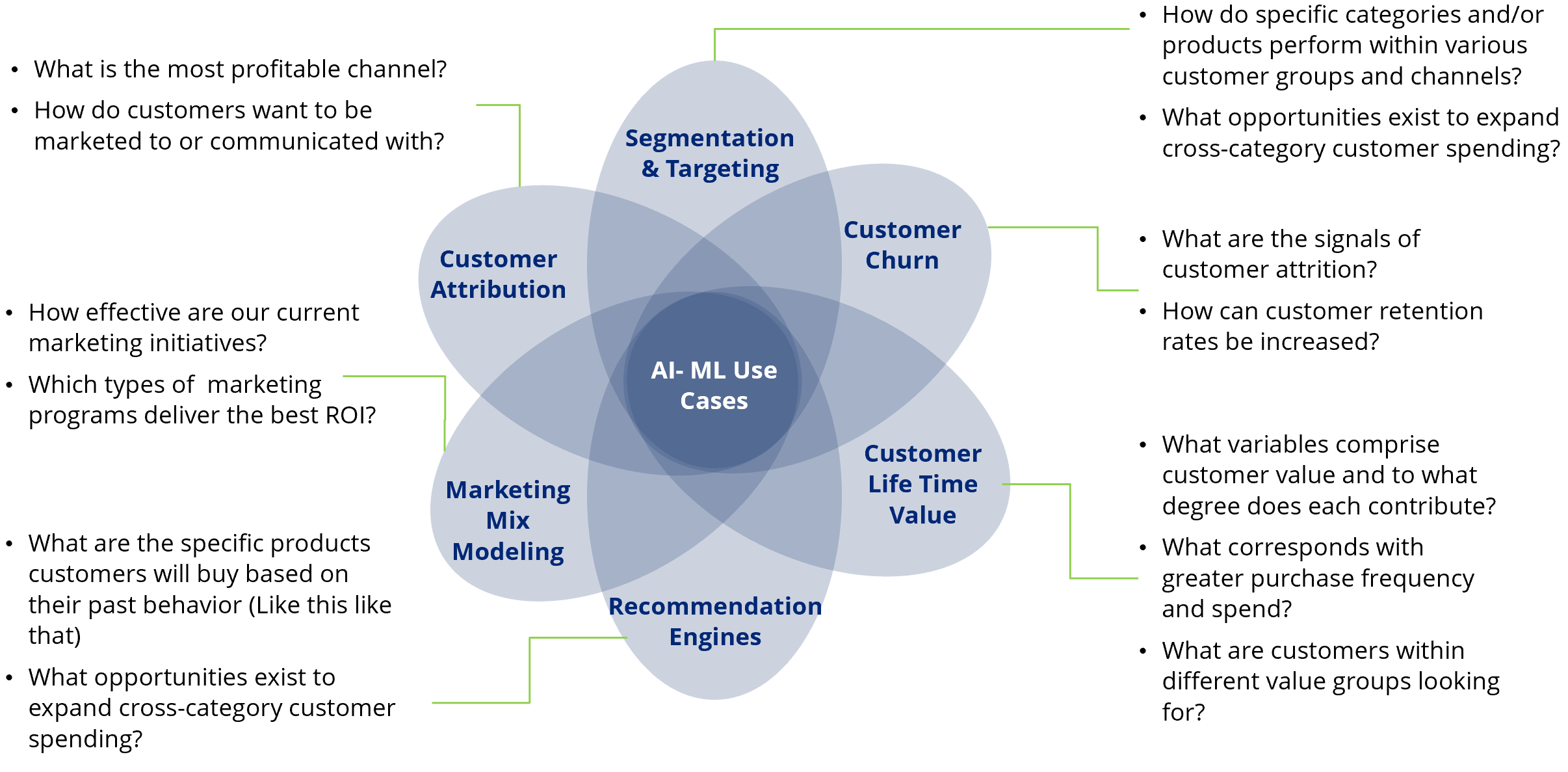
Segmentation also doesn’t have to be static. Dynamic segmentation is an application of AI that takes into account the fact that customers’ behaviors are rarely fixed or unchanging, and that people can take on different personas at different times for different reasons.

2.3.2 Machine Learning in Marketing

Machine learning can be used to predict behavior such as affinity for a given product or churn probability. However, this approach becomes slightly more challenging if you want to cluster similar customers, when there is no “ground truth”.

K-means and hierarchical aggregation are currently the most widely used algorithms to cluster datasets without human supervision. Each point (customers in the dataset) is assigned to a class. Leaving behind the limited perception of the human researcher, including hidden biases and presuppositions, these algorithms can spot the most obscure and surprising and least obvious clusters within the dataset.

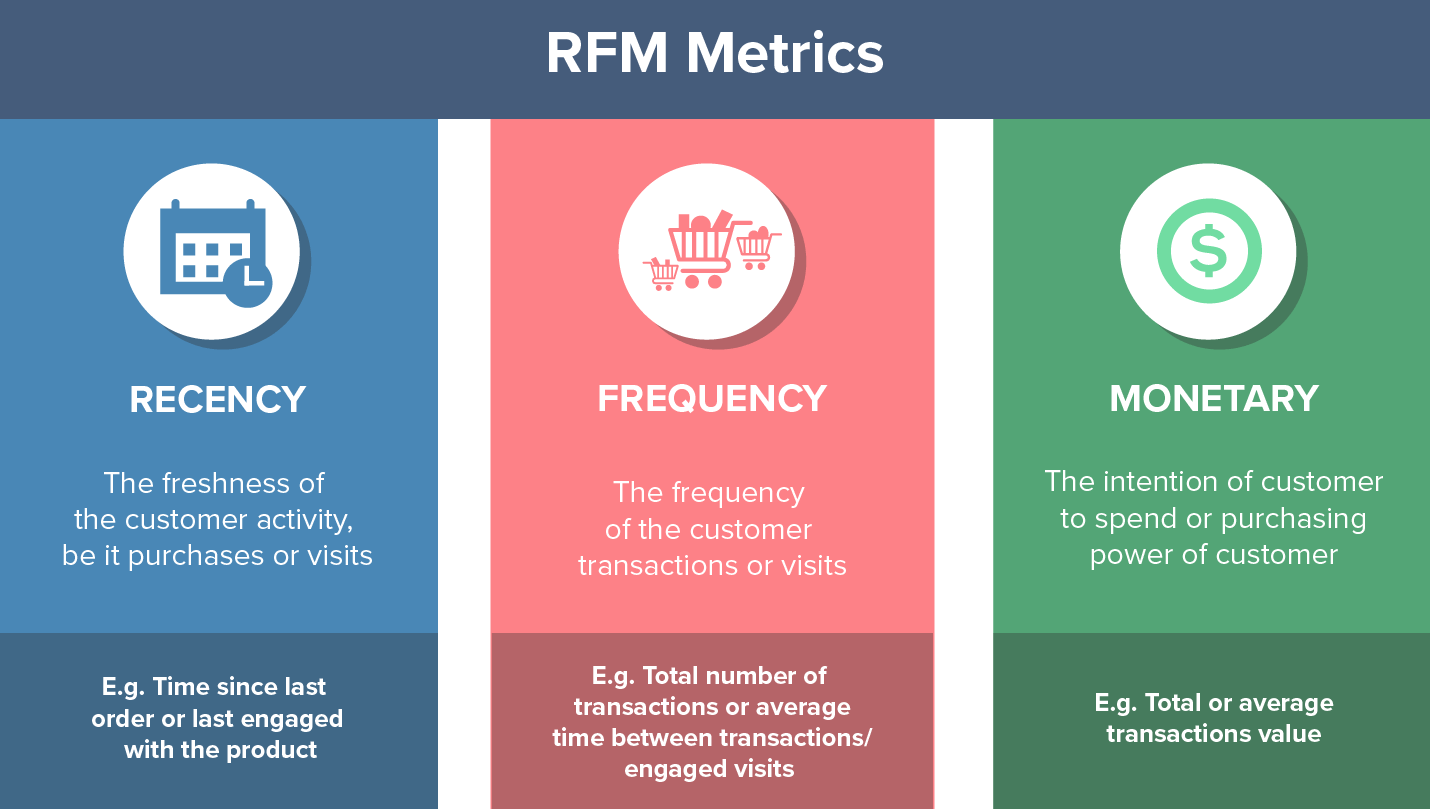
Machine learning brings lots of benefits for large organizations with data-oriented culture, which process great amounts of data and are able to leverage the system.



*Figure 2.4. AI & ML in marketing.*

**2.4** **RFM Analysis**

RFM (recency, frequency, monetary) analysis is a marketing technique used to determine quantitatively which customers are the best ones by examining how recently a customer has purchased (recency), how often they purchase (frequency), and how much the customer spends (monetary).



*Figure 2.5. RFM Metrics.*

There are 11 segments recommended for RFM:

|  |  |  |
| --- | --- | --- |
| **Customer Segment** | **Activity** | **Actionable Tip** |
| Champions | Bought recently, buy often and spend the most! | Reward them. Can be early adopters for new products. Will promote your brand. |
| Loyal Customers | Spend good money with us often. Responsive to promotions. | Upsell higher value products. Ask for reviews. Engage them. |
| Potential Loyalist | Recent customers, but spent a good amount and bought more than once. | Offer membership / loyalty program, recommend other products. |
| Recent Customers | Bought most recently, but not often. | Provide on-boarding support, give them early success, start building relationship. |
| Promising | Recent shoppers, but haven’t spent much. | Create brand awareness, offer free trials |
| Customers Needing Attention | Above average recency, frequency and monetary values. May not have bought very recently though. | Make limited time offers, recommend based on past purchases. Reactivate them. |
| About To Sleep | Below average recency, frequency and monetary values. Will lose them if not reactivated. | Share valuable resources, recommend popular products / renewals at discount, reconnect with them. |
| At Risk | Spent big money and purchased often. But long time ago. Need to bring them back! | Send personalized emails to reconnect, offer renewals, provide helpful resources. |
| Can’t Lose Them | Made biggest purchases, and often. But haven’t returned for a long time. | Win them back via renewals or newer products, don’t lose them to competition, talk to them. |
| Hibernating | Last purchase was long back, low spenders and low number of orders. | Offer other relevant products and special discounts. Recreate brand value. |
| Lost | Lowest recency, frequency and monetary scores. | Revive interest with reach out campaign, ignore otherwise. |

*Table 2. Recommended segments for RFM.*

RFM considers recency, frequency and monetary values for each customer. Combines them, and then groups them into different customer segments for easy recall and campaign targeting. RFM analysis is super useful in understanding responsiveness of your customers and for segmentation driven database marketing.

**2.5** **Clustering Algorithm**

There are 3 types of clustering algorithms we will consider for our project:

- K-mean

- Fuzzy C-mean

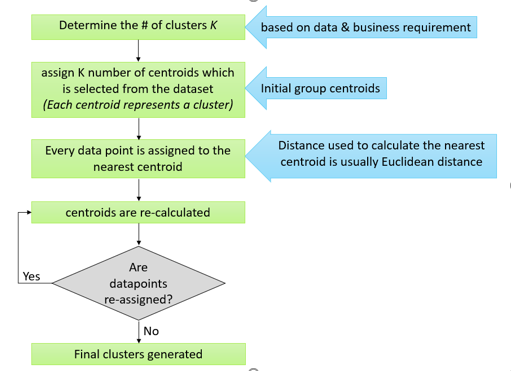
- Density-based

|  |  |  |  |
| --- | --- | --- | --- |
| Algorithm | K-mean | Fuzzy c-means | Density-based |
| Definition | - K-means clustering aims to partition *n* observation k-clusters in which each observation belongs to the cluster with the nearest mean, serving as a prototype of the cluster | - Fuzzy c-means is a method of clustering which allows one piece of data to belong to two or more clusters | - Detecting areas where points are concentrated and where they are separated by areas that are empty or sparse.  - Points that are not part of a cluster are labeled as noise. |
| Pros and Cons | - Suitable in a large dataset.  - Easy to interpret  - No optimal set of clusters.  - Accuracy  - Lack of consistency | - Allowed data point to be in multiple clusters.  - More natural representation  - Need to be define number of clusters | - Discovers clusters of arbitrary shapes.  - Handles noise.  - Cannot handle varying densities.  - Sensitive to parameters. |
| Use case | - Customer segmentation.  - Document clustering | - Marketing  - Image analysis |  |

*Table 3. Comparison between algorithm.*

2.5.1 K-means Clustering

K-means algorithm is an iterative algorithm that tries to partition the dataset into K-pre-defined distinct non-overlapping subgroups (clusters) where each data point belongs to only one group. It tries to make the inter-cluster data points as similar as possible while also keeping the clusters as different (far) as possible. It assigns data points to a cluster such that the sum of the squared distance between the data points and the cluster’s centroid (arithmetic mean of all the data points that belong to that cluster) is at the minimum. The less variation we have within clusters, the more homogeneous (similar) the data points are within the same cluster.

**

*Figure 2.6. K-means algorithm.*

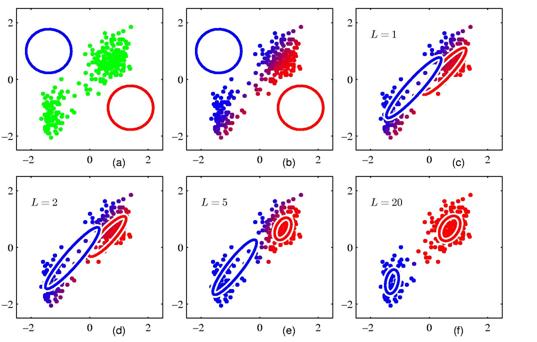
The way k-means algorithm works is as follow:

* Specify number of clusters K.
* Initialize centroids by first shuffling the dataset and then randomly selecting K data points for the centroids without replacement.
* Keep iterating until there is no change to the centroids. i.e. assignment of data points to clusters isn’t changing.
* Compute the sum of the squared distance between data points and all centroids.
* Assign each data point to the closest cluster (centroid).
* Compute the centroids for the clusters by taking the average of the all data points that belong to each cluster.

The approach k-means follows to solve the problem is call Expectation-Maximization.

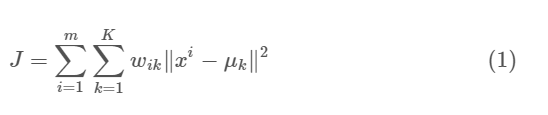
* The EM algorithm is used to find (local) maximum likelihood parameters of a statistical model in cases where the equations cannot be solved directly.

The E-step is assigning the data points to the closest cluster. The M-step is computing the centroid of each cluster.



*Figure 2.7. K-means visualization.*

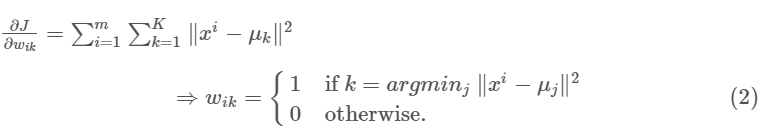
The objective function is:



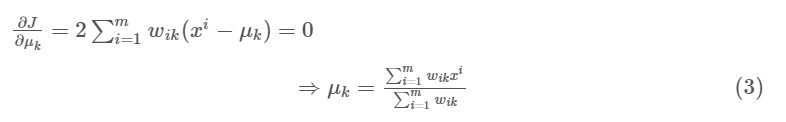
* + Wik = 1 for data point xi if it belongs to cluster k
  + otherwise, wik = 0. Also, μk is the centroid of xi’s cluster.

This formula is a minimization problem of two parts. We first minimize J w.r.t wik and treat μk fixed. Then we minimize J w.r.t. μk and treat wik fixed.

We differentiate J w.r.t. wik first and update cluster assignments (E-step). Then we differentiate J w.r.t. μk and recompute the centroids after the cluster assignments from previous step (M-step). Therefore, E-step is:



M-step is used to translates to recomputing the centroid of each cluster to reflect the new assignments.



2.5.2 Optimal K using Elbow Method

The Elbow method is a heuristic method of interpretation and validation of consistency within cluster analysis designed to help finding the appropriate number of clusters in a dataset

One technique to choose the best k is called the elbow method. This method uses within-group homogeneity or within-group heterogeneity to evaluate the variability.

In other words, we are interested in the percentage of the variance explained by each cluster. We can expect the variability to increase with the number of clusters, alternatively, heterogeneity decreases.

The challenge is to find the “k” that is beyond the diminishing returns. Adding a new cluster does not improve the variability in the data because very few information is left to explain.

Step to find the optimal “k”:

* Construct a function to compute the total within clusters sum of squares.
* Run the algorithm times.
* Create a data frame with the results of the algorithm.
* Plot the results.

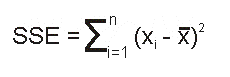
2.5.3 Internal Accuracy Measure

* Internal cluster validation, which uses the internal information of the clustering process to evaluate the goodness of a clustering structure without reference to external information. It can be also used for estimating the number of clusters and the appropriate clustering algorithm without any external data.

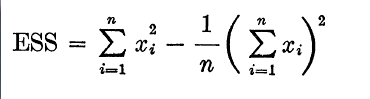
Our project considers 2 methods to evaluate compactness of a cluster:

* Sum of square error: SSE is the sum of the squared differences between each observation and its group's mean. It can be used as a measure of variation within a cluster. If all cases within a cluster are identical the SSE would then be equal to 0.

Formula for Sum of square error:



* + n is the number of observations
  + xi is the value of the ith observation
  + 0 is the mean of all the observations



At each stage of cluster analyzing, the total SSE is minimized with

SSE (total) = SSE (1) + SSE (2) + SSE (3) + SSE (4) .... + SSE (n).

At the initial stage when each case is its own cluster this of course will be 0.

2.5.4 External Accuracy Measure

External cluster validation, which consists in comparing the results of a cluster analysis to an externally known result, such as externally provided class labels. It measures the extent to which cluster labels match externally supplied class labels. Since we know the “true” cluster number in advance, this approach is mainly used for selecting the right clustering algorithm for a specific data set.

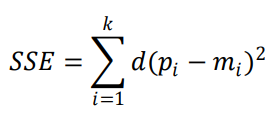
* Rand Index or Rand measures is a measure of similarity between two data clustering. A form of the Rand index may be defined that is adjusted for the chance grouping of elements, this is the adjusted Rand index. From a mathematical standpoint, Rand index is related to the accuracy, but is applicable even when class labels are not used.

2.5.5 Enhance K-Means Algorithm with Davies-Bouldin Index

Davies- Bouldin Index (DBI) is one methods used for measure validity cluster on something method clustering. Measurement with Davies - Bouldin Index is maximizing inter-cluster distance and on same time try for minimize distance between point in a cluster. If maximum inter-cluster distance, means similarity characteristics between each cluster a bit so difference inter-clusters visible more clear. If minimal intra-cluster distance means each object in the cluster have level similarity the characteristics of the high [1]. Clustering results obtained from proposed determining centroid then in evaluation with DBI method. So that could is known correlation from determining centroid method based on Sum of Squared Error to enhancement quality cluster based on the value of DBI obtained.

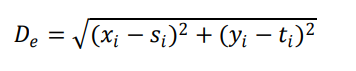
The step of proposed method for determining initial centroid in this paper are following:

1. Determine and input number of cluster
2. Determine number of centroid will be tested
3. Choose centroid with randomly
4. Calculate Sum of Squared Error (SSE) value centroid respectively



And then get centroid with minimum SSE from all centroid have been calculated. Back to step 2 until number of centroid is tested equal with input number of centroid. Centroid with minimum Sum of Square Error is used as initial centroid in clustering step.

1. Calculate distance between data and centroid with Euclidean distance formula.



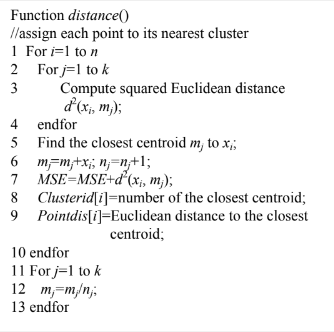
1. Classify data with minimum distance from step 5
2. Calculate and get new centroid based on mean value from membership of cluster
3. If new centroid value equal input centroid then clustering is stopped, else back to step 5.

In Enhancement Clustering Evaluation Result of Davies-Bouldin Index with Determining Initial Centroid of K-Means Algorithm [2], based on the result on the three datasets: Birth and Death Rates, Wholesale Customers and Seeds Dataset has shown that the method of determining centroid based on the minimum Sum of Squared Error (SSE) value can improve in terms of improving the clustering quality and show better clustering results compared to the method of conventional determining centroid by the difference of Davies-Bouldin Index value of 0.0378 or 3.78%. And then the comparison result value of previous research with current research is 0.0301 or 3.01 %. Based on this research, future work is to compare result with used categorical data, image data or determining centroid based on SSE with another clustering and then to compare with another evaluation clustering method.

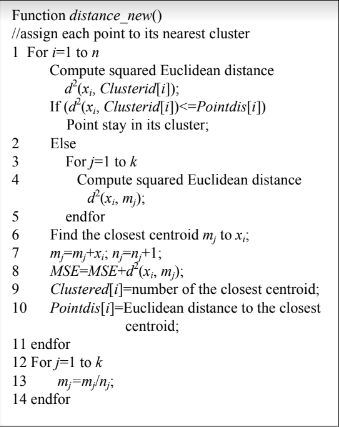
2.5.6 Enhance K-Means Algorithm By Optimizing Time

In each iteration, the k-means algorithm computes the distances between data point and all centers, this is computationally very expensive especially for huge datasets. Therefore, it came up with an idea that we can keep the distance to the nearest cluster for each data point. At the next iteration, we compute the distance to the previous nearest cluster. If the new distance is less than or equal to the previous distance, the point stays in its cluster, and there is no need to compute its distances to the other cluster centers. This saves the time required to compute distances to k−1 cluster centers[3].

By applying 2 functions: distance() and distance\_new(), we execute function distance() two times, while fuction distance\_new() is executed the reminder of iteration.



*Figure 2.x. Function distance().*



*Figure 2.x. Function distance\_new()*

In the proposed enhanced k-means algorithm, to obtain initial clusters, this process requires O(nk). Here, some points remain in its cluster, the others move to another cluster. If the point stays in its cluster this require O(1), otherwise require O(k). If we suppose that half points move from their clusters, this requires O(nk/2), since the algorithm converges to local minimum, the number of points moved from their clusters decreases in each iteration. So we expect the total cost is

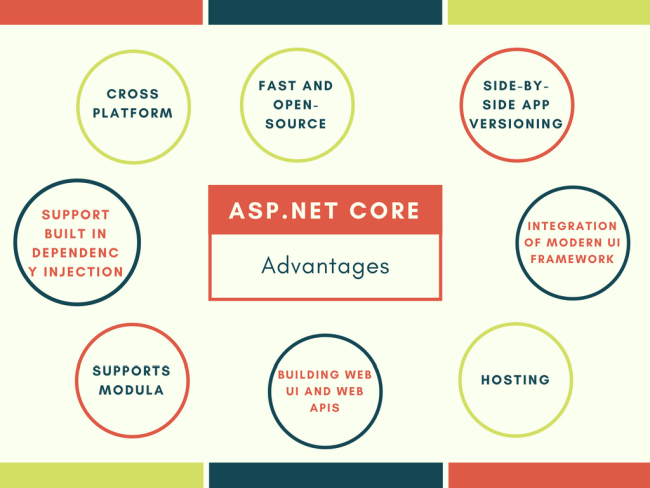


Even for large number of iterations,

 is much less than nkl. So the cost of using enhanced k-means algorithm approximately is O(nk), not O(nkl).

**2.6** **ASP.NET Core**

ASP.NET Core is the new web framework from Microsoft. It has been redesigned from the ground up to be fast, flexible, modern, and work across different platforms. Moving forward, ASP.NET Core is the framework that can be used for web development with .NET. If you have any experience with MVC or Web API over the last few years, you will notice some familiar features. At the end this tutorial, you will have everything you need to start using ASP.NET Core and write an application that can create, edit, and view data from a database.



*Figure 2.8. ASP.NET Core advantages.*

**2.7** **Front-end Technologies**

2.7.1 Bootstrap



Bootstrap is an open-source Javascript framework developed by the team at Twitter. It is a combination of HTML, CSS, and Javascript code designed to help build user interface components. Bootstrap was also programmed to support both HTML5 and CSS3. Also it is called Front-end-framework.

Bootstrap is a free collection of tools for creating a websites and web applications.

It contains HTML and CSS-based design templates for typography, forms, buttons, navigation and other interface components, as well as optional JavaScript extensions.

2.7.2 HTML



HTML is a computer language devised to allow website creation. These websites can then be viewed by anyone else connected to the Internet. It is relatively easy to learn, with the basics being accessible to most people in one sitting; and quite powerful in what it allows you to create. It is constantly undergoing revision and evolution to meet the demands and requirements of the growing Internet audience under the direction of the W3C, the organization charged with designing and maintaining the language.

2.7.3 MockFlow



MockFlow is a cloud-based wireframe mockup software solution designed for website and software developers, designers, UX professionals, and marketers. It lets them perform tasks, design user-interface prototypes for websites and mobile apps, and collaborate with projects in real-time.

Wireframing is a valuable tool for people who work on website design and mobile applications. As a result, MockFlow is the ideal online software solution not just for website designers, but also for decision-makers and stakeholders who want to put their ideas into a much more comprehensible mockup or wireframe to show to designers and vice versa.

As a wireframing software solution, MockFlow allows users to create a mockup or a skeleton of a website or mobile application they want to develop. The system provides a basic layout of the application, including its features and functions, without having to build the actual website. Developers can use this platform to show their clients the basic framework and to get a final approval or to make the necessary adjustments otherwise.

**2.8** **Back-end Technologies**

2.8.1 C#



C# is a programming language on computer science is a language that is used to write software programs. It was developed and launched by Microsoft in 2001. C# is a simple, modern, and object-oriented language that provides modern day developers flexibility and features to build software that will not only work today but will be applicable for years in the future.

**Chapter 3**

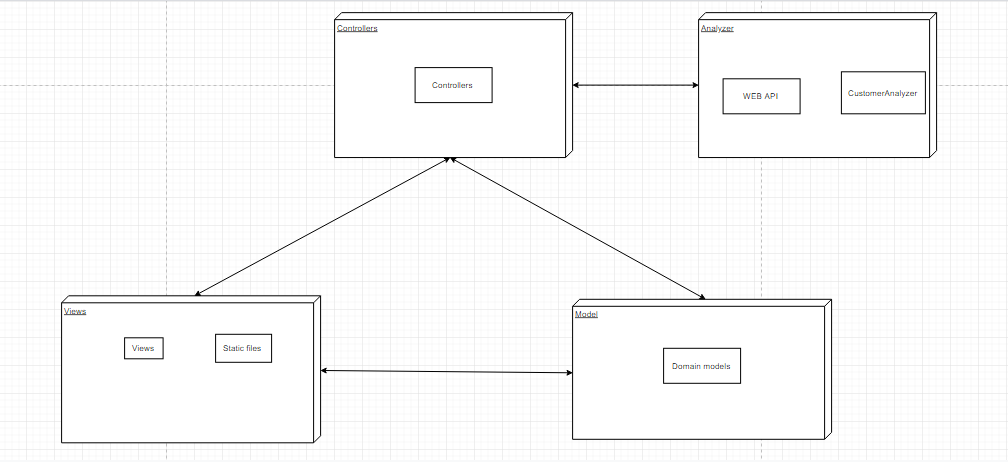
**APPLICATION ANALYSIS AND DESIGN**

**3.1** **Requirements**

Functional requirements:

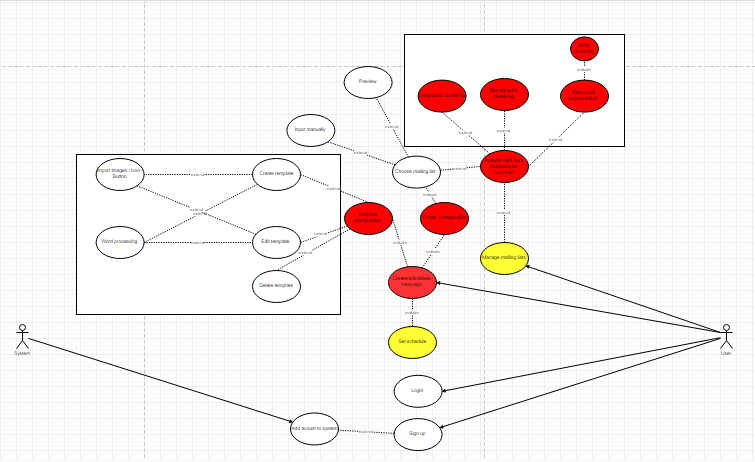
* Users shall be able to upload a PDF file with the size smaller than 10MB to the server.
* Users shall be able to select multiple grammar structures that they want to annotate on the document.
* Users shall be able to view the annotated file after processed on the webpage.
* Users shall be able to download the annotated PDF file from the server to the local machine.
* Users shall be able to check the tense of a sentence.

**3.2** **Architecture**

****

*Figure 2.9. Architecture.*

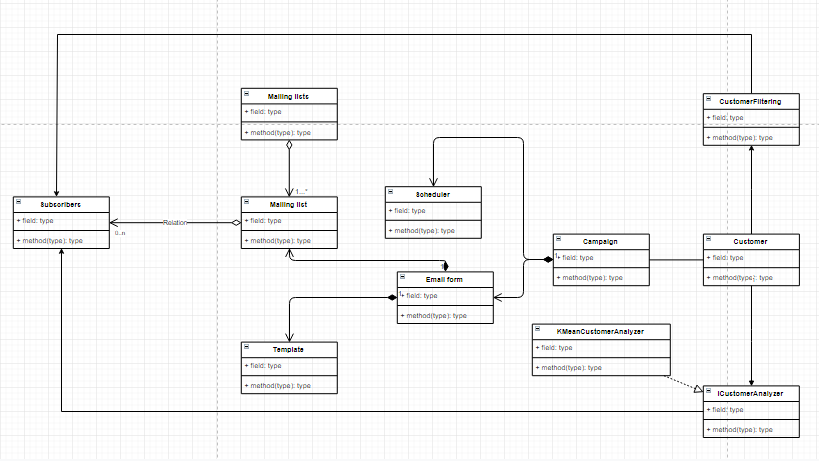
**3.3** **Use-case Diagram**



*Figure 3.1. Use case diagram.*

Description: …

**3.4** **Class Diagram**



*Figure 3.2. Class diagram.*

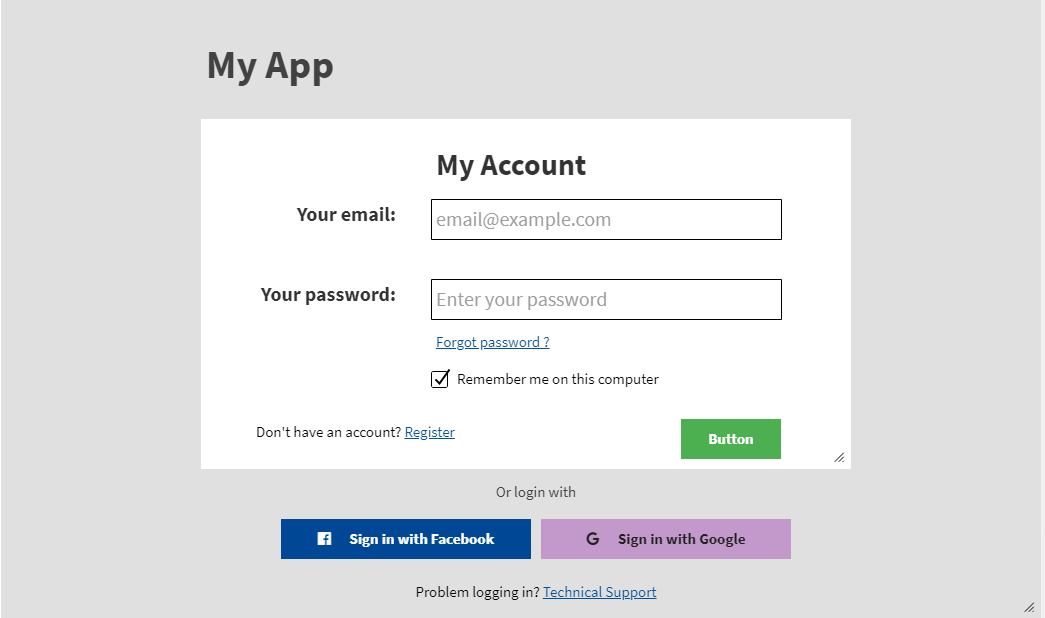
Description: …

**3.5** **Activity Diagram**

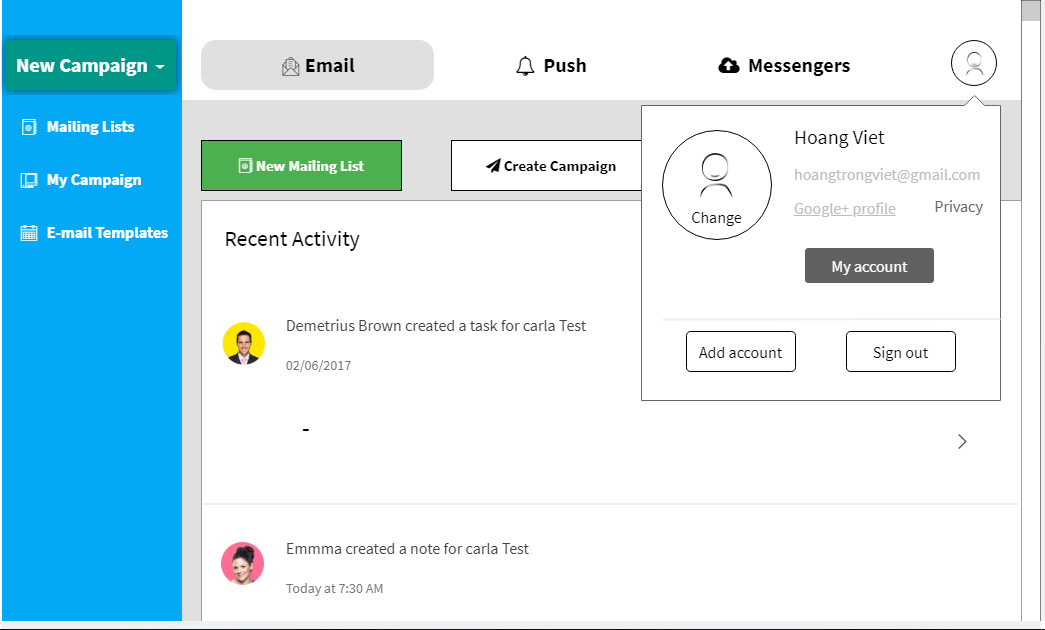
*Figure 3.3. Activity diagram.*

Description: …

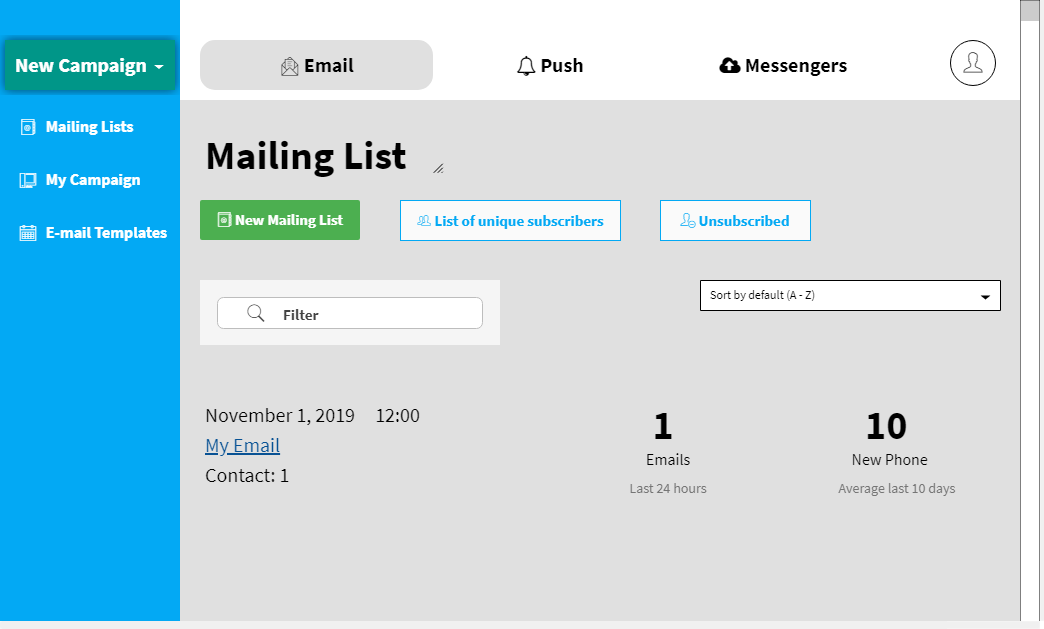
**3.6** **UI Design Using WireframePro**

****

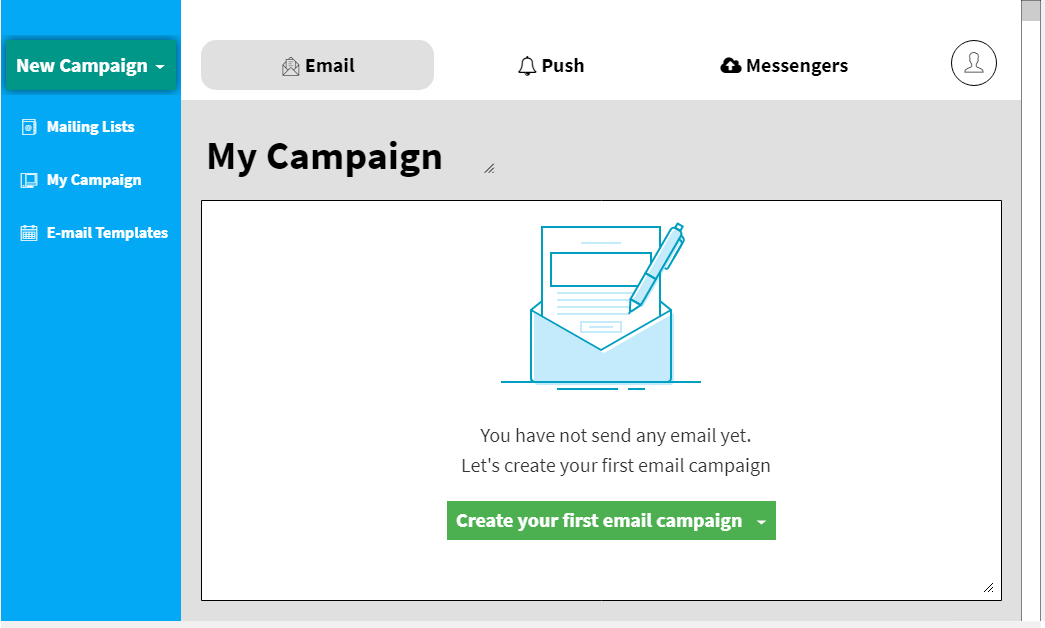
*Figure 3.4. Login screen.*



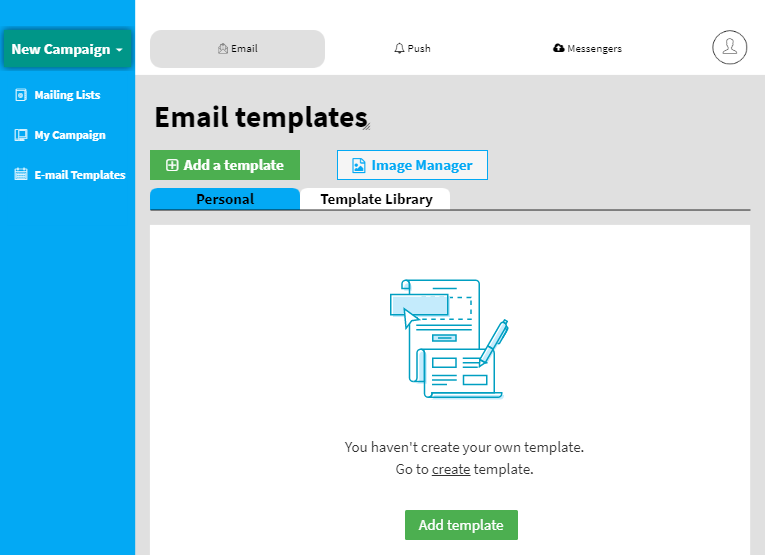
*Figure 3.5. Dashboard screen.*

**

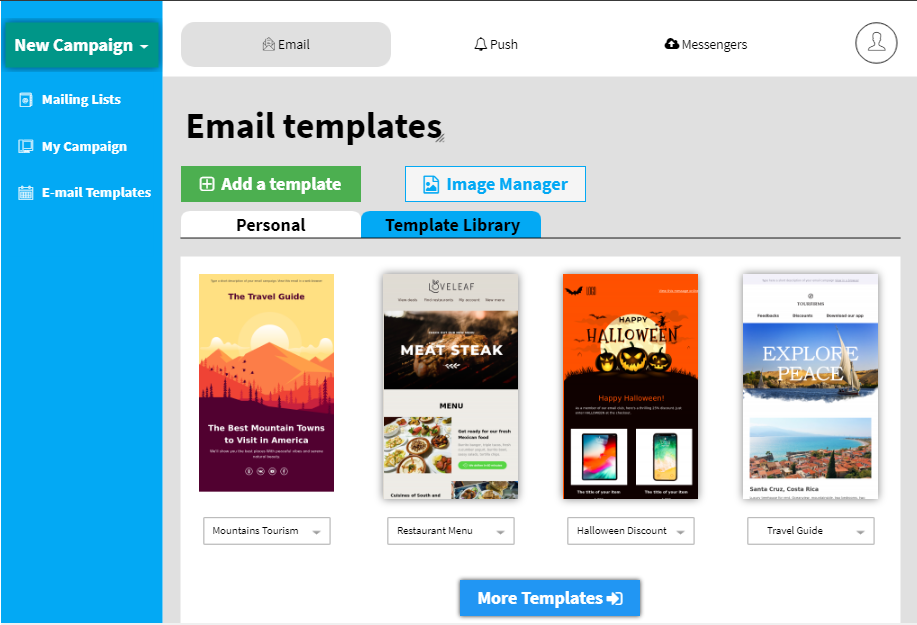
*Figure 3.6. Mailing list screen.*

**

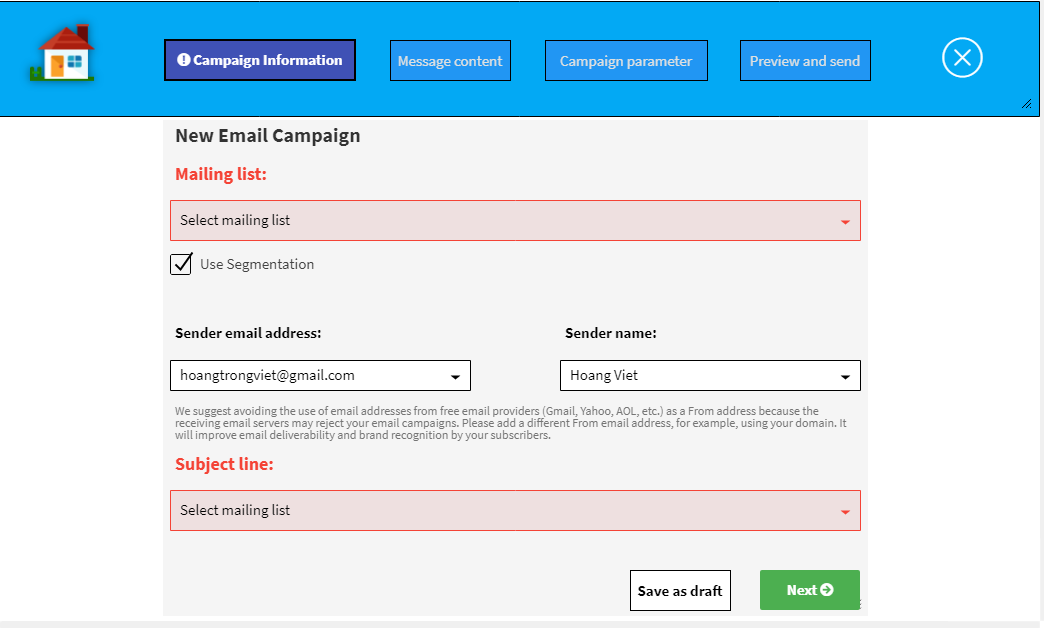
*Figure 3.7. My campaign screen.*

**

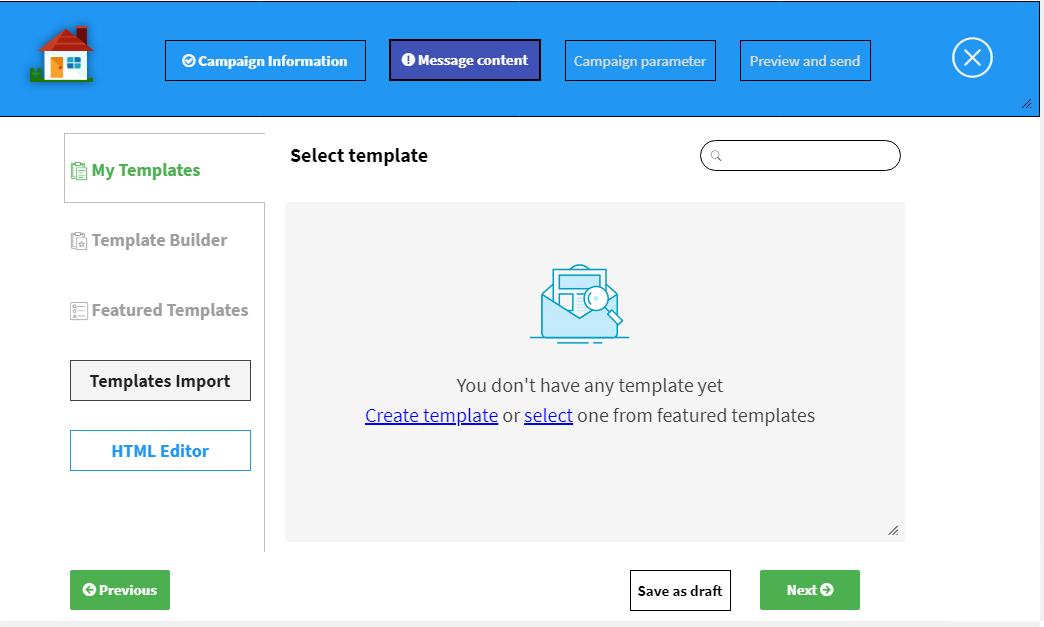
*Figure 3.8. Add Template screen.*

**

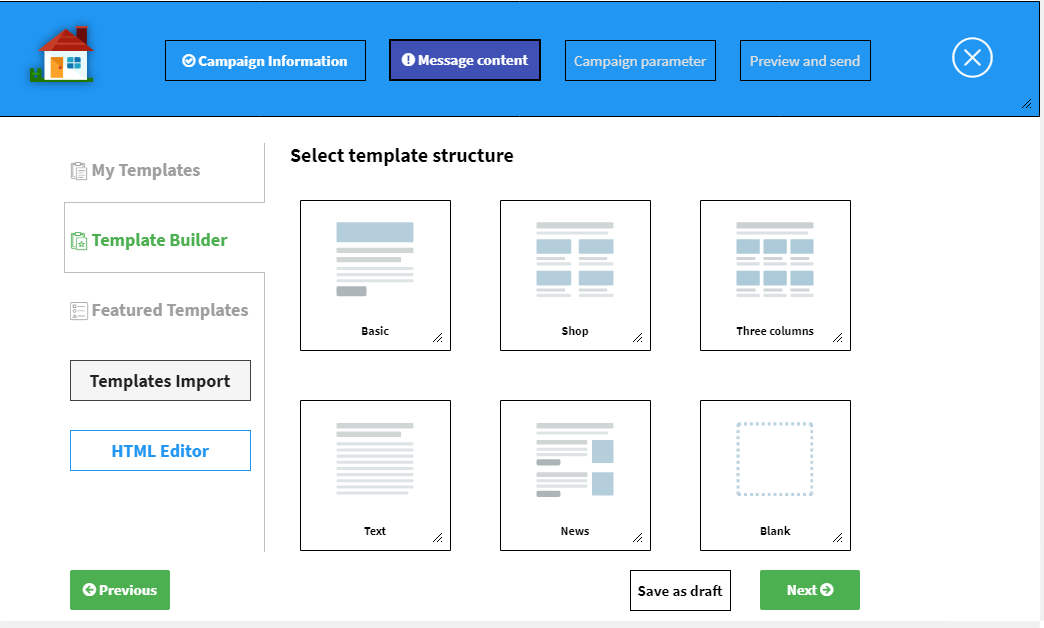
*Figure 3.9. Template Library screen.*

**

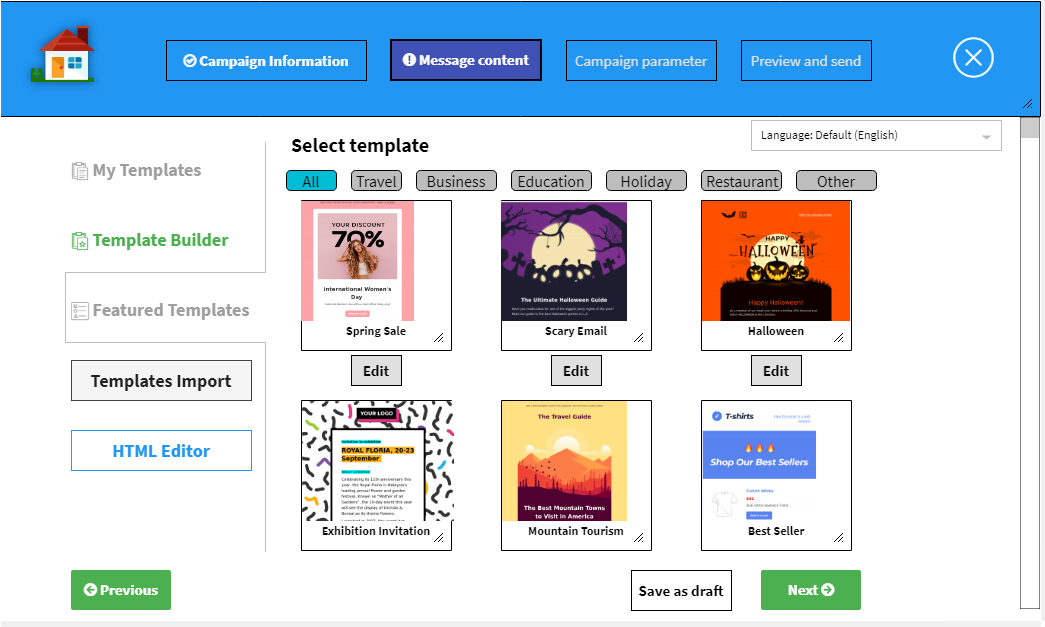
*Figure 4.1. Create campaign step 1.*

**

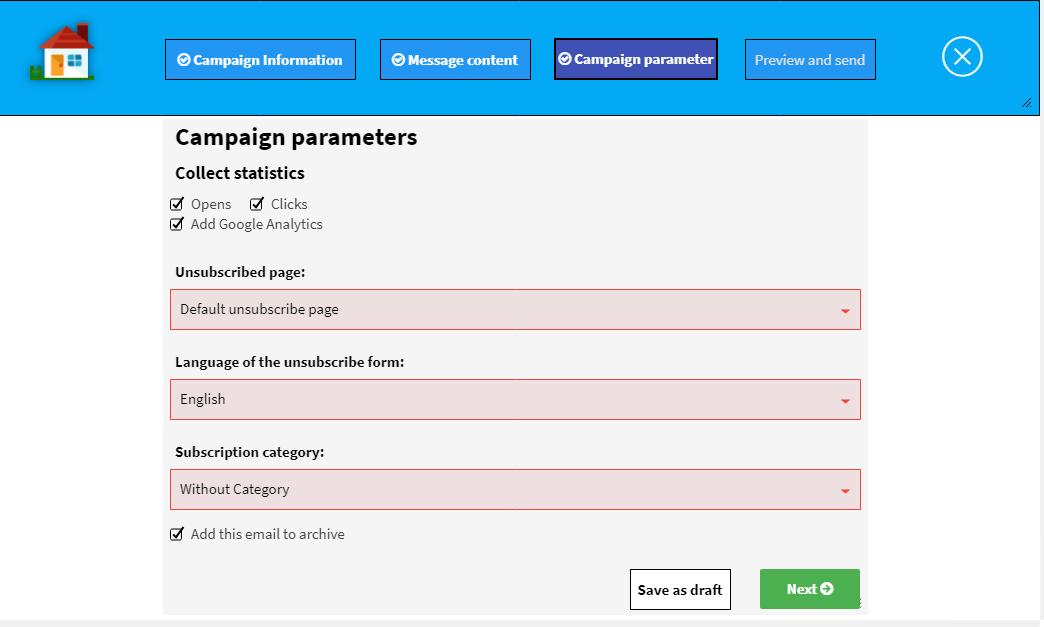
*Figure 4.2. Create campaign step 2.*

**

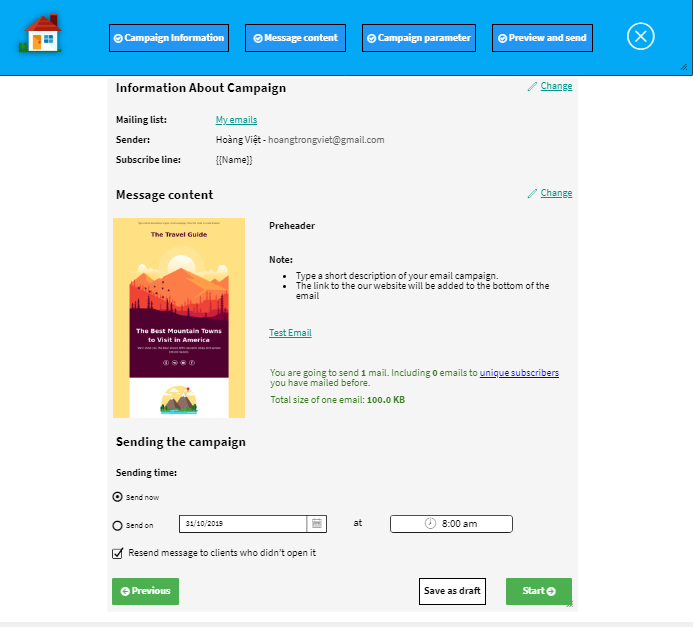
*Figure 4.3. Create campaign step 3.*

**

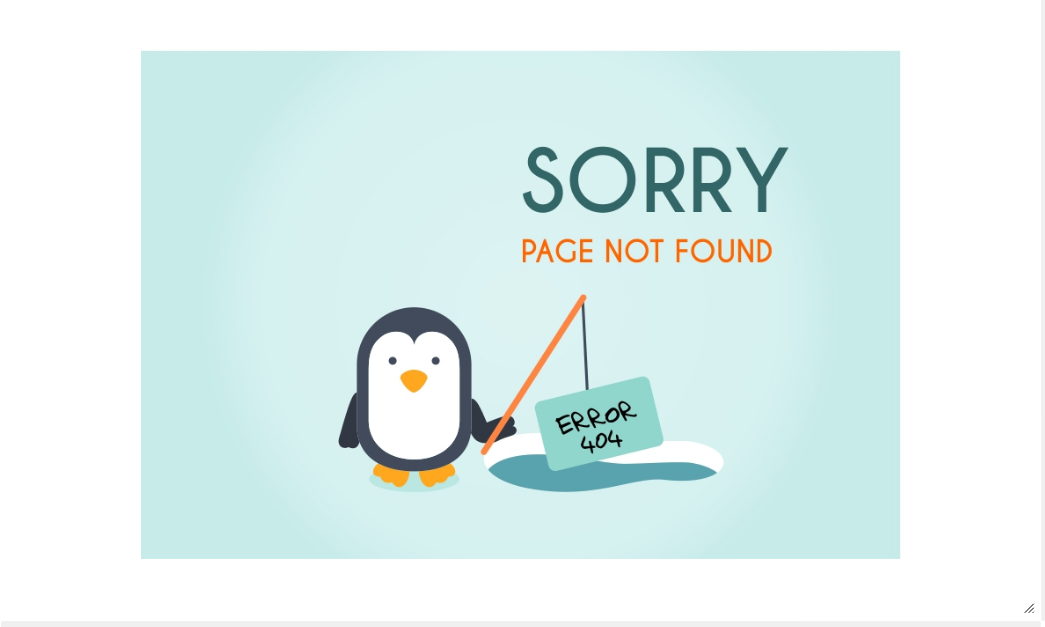
*Figure 4.4. Create campaign step 4.*

**

*Figure 4.5. Create campaign step 5.*

**

*Figure 4.6. Create campaign step 6.*

**

*Figure 4.7. Error page.*

**Chapter 4**

**MARKETING APPLICATION IMPLEMENTATION**

**Chapter 5** **EVALUATION**

**Chapter 6** **CONCLUSION AND FUTURE WORKS**

**6.1** **Conclusion**

**6.2** **Future works**

**REFERENCES**

1. Thakare, Y.S. & Bagal, S.B. 2015. Performance evaluation of k-means clustering algorithm with various distance metrics. International Journal of Computer Applications 110(11): 12-16.
2. Bernad Jumadi Dehotman Sitompul et al 2019 J. Phys.: Conf. Ser. 1235 012015. Enhancement Clustering Evaluation Result of Davies-Bouldin Index with Determining Initial Centroid of K-Means Algorithm
3. Fahim et al. / J Zhejiang Univ SCIENCE A 2006 7(10):1626-1633. An efficient enhanced k-means clustering algorithm.