**Project Name:** Marketing Optimization Project

Contents

[Executive Summary 3](#_Toc428795880)

[Overview 4](#_Toc428795881)

[Goal 4](#_Toc428795882)

[Scope 4](#_Toc428795883)

[Team 4](#_Toc428795884)

[Due Date 4](#_Toc428795885)

[Project Plan 5](#_Toc428795886)

[Phase 1 – Methodology and Model 5](#_Toc428795887)

[1. Customer Recognition 6](#_Toc428795888)

[Challenges 6](#_Toc428795889)

[2. Data Collection 6](#_Toc428795890)

[Challenges 6](#_Toc428795891)

[3. Data Mining 7](#_Toc428795892)

[Challenges 7](#_Toc428795893)

[4. Product Recommendation 7](#_Toc428795894)

[Challenges 7](#_Toc428795895)

[Phase 2 – Troubleshooting 8](#_Toc428795896)

[1. Customer Recognition 8](#_Toc428795897)

[Lessons learned 8](#_Toc428795898)

[2. Data Collection 8](#_Toc428795899)

[Lessons learned 8](#_Toc428795900)

[3. Data Mining 10](#_Toc428795901)

[Lessons learned 10](#_Toc428795902)

[4. Product Recommendation 11](#_Toc428795903)

[Lessons learned 11](#_Toc428795904)

[Phase 3 – Beta & Pilot 13](#_Toc428795905)

[Module 1 - Offline Setup of Classification Model 13](#_Toc428795906)

[Module 2 - Data Collection & Mining 13](#_Toc428795907)

[Module 3 - Personal Content Trigger 14](#_Toc428795908)

[Process State Transition Diagram 15](#_Toc428795909)

[Conclusions & Next steps 17](#_Toc428795910)

[References 18](#_Toc428795911)

# Executive Summary

This project is focused on developing a new digital signage technology using data mining. The whole solution consists of three modules: offline setup of the classification model (module 1), data collection & mining (module 2), and personal content trigger (module 3). A prototype has been built up and the live demo works properly and smoothly. Further developments are needed, and some suggestions are given in the end of this document.

In order to operate data mining, a large volume of homologous data is needed as a training dataset. After pre-processing, training set will be sent to train the data model. This is module 1 and this module can be done offline in advance. Module 2 is the main body of the data mining algorithm. For the first step, the user’s social network information will be fetched, and then the data will go through the same pre-processing as in module 1. Lastly, the machine will give its prediction on the processed data. Module 2 is real-time, which consumes only 1 second, observed from experiments. Having the data mining result, module 3 is using this result to display the corresponding contents on certain players. HTTP polling is set up between server and players. Once any detection is received, the content on the player could be changed real-time.

# Overview

Rapid developments in information technology, coupled with the generation of large amounts of data, has quickly driven big data initiatives and technologies to become the next mega trend that industries try to leverage as a competitive advantage.

The above technologies have enabled a new type of marketing, known as customer-centric marketing[[1]](#footnote-1). In traditional marketing, products, services and channels are driven based on an “average” customer profile. In contrast, customer-centric marketing allows companies to look past an “average” customer, and treat each customer as unique. However, this requires companies to really understand the needs and wants of their customers, which “big data” initiatives hope to unleash.

# Goal

The goal of this project is to create a technology that would enable a financial institution, such as the Bank of Montreal (BMO), to capture, understand and predict their customers’ wants and needs with regard to their product offerings.

# Scope

The scope of the project is limited to the following:

* Develop a technology that can be used in a financial institution (BMO) application
* Integrate with current digital signage technology at BMO
* Data mining to include both internal and external sources to the bank (i.e. social media, bank databases, etc.)
  + Use data sources that are available (i.e. mostly external sources, since internal bank sources will require bank approval)
* Limit to current existing BMO customers only

# Team

* Xulai C – Technology developer
* Bill L – Solutions Architect
* Daniel J – Developer

# Due Date

End of August 2015

# Project Plan



Fig.1 - Project Plan

# Phase 1 – Methodology and Model

This section develops the methodology and model to be used and prototype in the next phase.

This is a preliminary high level view of the project model:

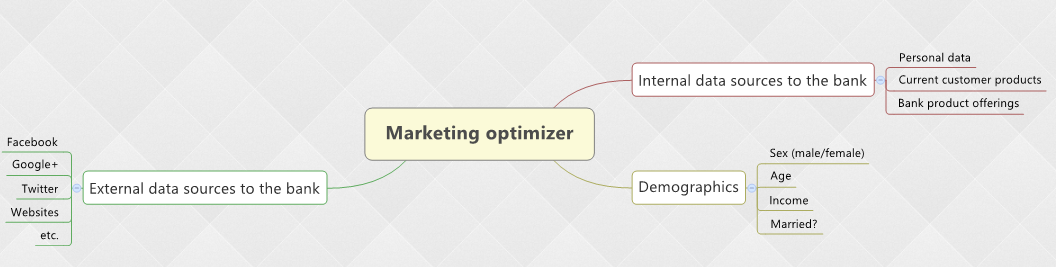


Fig.2 - Project Plan

Analyzing and mining customers’ needs can be divided into 4 stages.

Customer Recognition

Data Collection

Data Mining

Product Recommendation

Fig.3 - Project Plan

## Customer Recognition

Once a customer enters the effective area of one digital player, we need to recognize him. Some of his information needs to be obtained in order to deliver further process.

### Challenges

1. Device Choice. We need to determine the device that detects and recognizes customers.
2. Limited Customer Information. Due to the technical restrictions and potential privacy policy, we need to clarify which personal identification we can obtain at stage 1, which will significantly influences the schema of next stages.

## Data Collection

Using the customer identifier we obtain in stage 1, we will collect their social media data and transform it into a processible format.

### Challenges

1. Account Access. How to access their social media account, extract the data we need with limited personal information we had in stage 1.
2. Data Transformation. How to transform the original raw data into a processible format.
3. Training dataset collection. In order to train the model, a large volume of training data is needed. Moreover, the training data need to be homologous with user data.

## Data Mining

Starting from the formatted social media data, we need to go through preprocessing, feature extraction & selection and classification successively to find out customers’ need. Moreover, due to access restriction to internal data sources from bank, unsupervised classification would be applied.

### Challenges

1. Public Dataset. Due to privacy policy of Facebook. It may be difficult to obtain proper user dataset to be training set.
2. Feature Extraction. How to extract informative features from unsorted data source.
3. Dimensionality. Usually social media covers a high volume of topics; dimensionality of features could be tremendous. Feature selection need to be carefully.
4. Unsupervised classification. How to analyze data without label (customer’s preference already know) will be another big challenge of project.

## Product Recommendation

Based on the results in stage 3, we determine the BMO products which certain customer may interest most, and then change the content of player to related information.

### Challenges

1. Product List. We need to determine BMO product list which BMO would like to promote.
2. Integration. How to apply what we have based on data mining to current digital signage technologies. How to smooth switchover of displaying content between regular one and triggered one.

# Phase 2 – Troubleshooting

## Customer Recognition

To identify customers, we choose to develop an app to ask for customers’ permission to their social network info. This app can be a new feature to be integrated with an existing app (like BMO’s app) or a start of our own app for digital signage. Android ID will be sent to server as an identifier of android device (customer).

### Lessons learned

* ANDROID\_ID  
  A 64-bit number (as a hex string) that is randomly generated when the user first sets up the device and should remain constant for the lifetime of the user's device. When a device has multiple users (available on certain devices running Android 4.2 or higher), each user appears as a completely separate device, so the ANDROID\_ID value is unique to each user. So we could use the unique android ID to recognize user. For the code, please refer to Android Project.
* Facebook node id

Facebook assigns a unique node id to each user, thus we can use Facebook node id as a customer identifier.

## Data Collection

The app will send customer Facebook access token to our server under permission. With token received, customers’ social network info will be fetched at server. Further processing will be operated at server then to get it ready for data mining. Some of Facebook node features are to be chosen, the choices will be determined based on the raw data.

### Lessons learned

* Facebook API  
  In order to retrieve user’s Facebook info, we need to fetch their data through Facebook API. Here are three basic Graph terms: nodes, fields, edges.

Nodes:

A node is an object returned from Graph like a user, page or event. If a page has an id of 1337 it can be accessed by the following endpoint:

*GET /1337*

Fields:

Fields reference specific properties of a Graph object. So a page object will have an ID, name and description among other things. You can specify which fields you'd like to be returned for a particular node using the fields modifier.

*GET /1337?fields=id,name,description*

Edges:

Data on Facebook is very relational. A page for example could have many timeline posts associated with it. This relation between two Graph objects is called an edge. We can grab the page timeline posts from a page with an ID of 1337 like so:

*GET /1337/posts*

Posts here is not a node since it doesn't represent a Graph object. Instead it represents a relation to other Graph objects. It is an edge because it connects a page Graph object and its posts.

* Nested requests syntax

To retrieve complicated data, Facebook API allows nested requests. The syntax for nested requests looks like this:

*/<node-id>?fields=<first-level>{<second-level>}*

Where <node-id> is the Graph object you want to retrieve, <first-level> is the name of the field or edge. If <first-level> is an edge, you can further specify that edge's fields or edges in <second-level>.

Edges can be embedded infinitely deep. We can construct nested request like this:

*GET /me?fields= id,name,*

*photos.limit(5){id,name,source},*

*likes.limit(3){id,name},*

*events{name,photos{source,likes{name}}}*

* Nodes may need
  + Posts(Status)

*/v2.3/me/posts?fields=message,description&access\_token=*

*/v2.3/me?fields=posts{message,description}&access\_token=*

* + Likes

*/v2.3/me/likes?fields=id,name,description,category&access\_token=*

*/v2.3/me?fields=likes{id,name,description,category}&access\_token=*

* + user\_about\_me

*/v2.3/me?fields=bio&access\_token=*

* + user\_actions.books

*/v2.3/me/books?fields=name,description&access\_token=*

*/v2.3/me?fields=books{name,description}&access\_token=*

* + user\_events

*/v2.3/me/events?fields=name,description&access\_token=*

*/v2.3/me?fields=events{name,description}&access\_token=*

* + user\_groups

*/v2.3/me/groups?fields=name,description&access\_token=*

*/v2.3/me?fields=groups{name,description}&access\_token=*

* + The complete graphic API url:

*/v2.3/me?fields=posts{message,description},*

*likes{id,name,description,category},*

*bio,*

*books{name,description},*

*events{name,description},*

*groups{name,description}*

*&access\_token=*

*https://graph.facebook.com/v2.3/me?fields=posts{message,description},likes{id,name,description,category},bio,books{name,description},events{name,description},groups{name,description}&access\_token=CAAUuQQjuVIEBAP1T11drd60gy0sjZASslEGWK2HRyBlt8WIfNrQALb9TpnrSFlbK7wiENmd91ErDMlYr35fvXoCjaVHWiVmcuv2PjqXZBbE9cWrUujeOQZBN1e0JQuHeRcK2isHESoMn3uaRlOiPt12px6GRknwMajQ439wYLDDsY5XkZCF7GXZA3ZATZBnrvbSzDhai3LLCdAWZBZB9geZC6ZBJihzWB7V5KZCFnp71bbOZBKgZDZD*

## Data Mining

To ensure accuracy of data mining, resources of training dataset and operating dataset need to be homologous. After exhaustive search online, we found no proper dataset available, so we composed a crawl script to crawl public Facebook group description and Facebook page description through Facebook Graph API. The reason why we choose group info and likes page info is because we believe these may imply people’s background and status. After fetching training set, we operate clustering (without labels) and analyze clustering results, then assign each group a proper label based on the results evaluation. Then send those labeled dataset to classifier to train the model. After training, classifier will be ready to work. All those setups of classification model will be finished offline prior to data mining. After that, customer content mining will be executed on our server.

### Lessons learned

* Product List

Text mining, also referred to as text data mining, roughly equivalent to text analytics, refers to the process of deriving high-quality information from text. High-quality information is typically derived through the devising of patterns and trends through means such as statistical pattern learning. Text mining usually involves the process of structuring the input text (usually parsing, along with the addition of some derived linguistic features and the removal of others, and subsequent insertion into a database), deriving patterns within the structured data, and finally evaluation and interpretation of the output.

Text analysis involves information retrieval, lexical analysis to study word frequency distributions, pattern recognition, tagging/annotation, information extraction, data mining techniques including link and association analysis, visualization, and predictive analytics. The overarching goal is, essentially, to turn text into data for analysis, via application of natural language processing (NLP) and analytical methods.

* Clustering

Cluster analysis or clustering is the task of grouping a set of objects in such a way that objects in the same group (called a cluster) are more similar (in some sense or another) to each other than to those in other groups (clusters). It is a main task of exploratory data mining, and a common technique for statistical data analysis, used in many fields, including machine learning, pattern recognition, image analysis, information retrieval, and bioinformatics.

## Product Recommendation

To deliver product recommendation, we need to determine target production list (called label in classification) first. These products should both be interests of BMO promos and have strong implication from certain social media content. Then server will be responsible to communicate with players to trigger players to play certain video clips when needed. This communication uses HTTP polling.

### Lessons learned

* Product List  
  Here is several BMO products which could benefit from social data mining, and also we already have videos for them. We may choose some of them to implement in pilot.
  + Travel Insurance

*Z:\BMO\01\_Standard\_Content\Standard\_Content\_2015\07\_July 15\July 14\ BMO\_S\_ENG\_07\_TravelInsurance\_071415*

* + Mortgage

*Z:\BMO\01\_Standard\_Content\Standard\_Content\_2015\05\_May 15\May 12\ BMO\_STD\_ENG\_03A\_Mortgage\_051215*

* + Credit Card (1.75% Cash Back)

*Z:\BMO\01\_Standard\_Content\Currently Playing National\ BMO\_STD\_ENG\_09\_WEMC\_060115\_v2*

* + Wealth Management

*Z:\BMO\01\_Standard\_Content\Currently Playing National\ BMO\_STD\_ENG\_02\_Halo\_041615*

* + Investor Line

*Z:\BMO\01\_Standard\_Content\Currently Playing National\ BMO\_STD\_ENG\_03\_aDProspectingDoggy\_121014*

*BMO\_STD\_ENG\_07\_aDInvestments\_121014*

*BMO\_STD\_ENG\_11\_aDQuantitiveAnalysis\_121014*

* + New Customer ($425 bonus)

*Z:\BMO\01\_Standard\_Content\Currently Playing National\ BMO\_STD\_ENG\_04\_EDB\_060115*

* + New start program

*Z:\BMO\01\_Standard\_Content\Standard\_Content\_2015\05\_May 15\May 18\ BMO\_STD\_ENG\_16\_NewStart\_051815*

* + BMO saving builder account

*Z:\BMO\01\_Standard\_Content\Standard\_Content\_2015\06\_June 15\June 2\06A\_SBA-Fanta\ BMO\_STD\_ENG\_06A\_SBA\_060215*

* + BizBasic Plan ($6/mon)

*Z:\BMO\01\_Standard\_Content\Currently Playing National\ BMO\_STD\_ENG\_08\_BizBasic\_060115*

* + Fresh Book (cloud accounting)

*Z:\BMO\01\_Standard\_Content\Standard\_Content\_2015\09\_september 15\Sept 1\17\_Fresh Books\ BMO\_STD\_ENG\_17\_FreshBooks\_090115*

* + eStatements

*Z:\BMO\01\_Standard\_Content\Standard\_Content\_2015\04\_April 15\April 1\04 eStatements\ BMO\_STD\_ENG\_04\_eStatement \_040115*

* HTTP polling

HTTP server push (also known as HTTP streaming) is a mechanism for sending unsolicited (asynchronous) data from a web server to a web browser. HTTP server push can be achieved through any of several mechanisms.

Polling is itself not a true push, it allows emulating a push mechanism under circumstances where a real push is not possible, such as sites with security policies that require rejection of incoming HTTP/S Requests.

# Phase 3 – Beta & Pilot

Basically, there are 3 modules in this solution: offline setup of classification model, data collection & mining, personal content trigger.

## Module 1 - Offline Setup of Classification Model

1. Fetch Training dataset for classifier via Facebook Graphic API search engine. Search key words are relevant to the products that BMO wants to promote. Mark each record with its search key as its temporary label.
2. Pre-process the raw data, transform it into a feature list and start clustering algorithms.
3. Based on the evaluation of clustering, finalize the labels of each data.
4. Now send those adjusted dataset to train classifier.

## Module 2 - Data Collection & Mining

This module can be done anytime, anywhere, not necessarily in financial institutions.



Fig. 4 - Data Collection & Mining

Here are the details of each step:

1. Facebook login. Log into Facebook and request for access token with certain permissions.
2. Facebook send back the access token to the android device.
3. Android device would retrieve customer’s Facebook node id, and send it along with access token and its own android id to Novramedia server.
4. All the info received from device will be stored in database; primary keys are android id and customer’s Facebook node id. And then send a get request to Facebook API with customer granted access token.
5. Fetch customer’s Facebook data at server.
6. Filter the unwanted info out and get it ready for data mining. Run *data\_mining.py* at server and store mining result into database.

## Module 3 - Personal Content Trigger

Once customer enters a branch and chooses to detect Bluetooth signal, this module starts.



Fig. 5 - Personal Content Trigger

Here are the details of each step:

1. IBeacons send out Bluetooth signal continuously.
2. Once android device captured the signal, distance will be calculated. If it’s under a threshold, the device will notify server to start a trigger.
3. Server and screen communicate via HTTP polling, and it is always on. When server receives trigger request from mobile devices, it will forward this message to player(screen) along with the name of a video clip determined by data mining result which has been stored before.
4. When player(screen) gets notification, it stops its current content and play the video clips right away.

## [Process State](https://www.google.ca/url?sa=t&rct=j&q=&esrc=s&source=web&cd=11&ved=0CD8QFjAKahUKEwjlz9jxwNHHAhWFQ5IKHSm6BaQ&url=http%3A%2F%2Fwww.cs.jhu.edu%2F~yairamir%2Fcs418%2Fos2%2Fsld004.htm&ei=5lTjVeXEDoWHyQSp9JagCg&usg=AFQjCNH7SrZjY8nGKeZsfQJ5ABuEjIeUDA&sig2=r-ABAj1SEclP6VLnPw3Meg&cad=rja) Transition Diagram

This application is integration of these three modules. Fig. 6 shows the work flow.



Fig. 6 - Process State

Code Reference:

Module 1:   
Fetching training dataset - php code  
Pre-processing - python code  
Clustering & relabeling - matlab code

Module 2:   
Fetching customer info - android java code & php code  
Pre-processing - python code  
Data analysis - python code & php code

Module 3:   
Customer request - android java code & php code   
Trigger(server side) - php code   
Trigger(player side) - java script code

## Demo Instruction

Before starting the demo, make sure that these files have been uploaded to our server (currently ther’re all there).

*novramedialab* server

path: */public\_html/nathan*

nathan.php beacon.php Log.php File.php connection.php

path: /public\_html/nathan/fb

facebook.php base\_facebook.php fb\_ca\_chain\_bundle

path: /public\_html/nathan/data\_mining\_py

data\_mining.py auxiliary\_functions.py svmutil.py svm.py trie.py

nltk(folder) libsvm-3.20(folder)

path: /public\_html/nathan/polling

message.php Log.php File.php connection.php util.php

player side

path: *C:\dmedia\navori\_trigger\backend*

main.js navori.js NavoriQLSdk.dll navoritrigger.exe

and all dependencies have been built, also main.js is **running**.

After all above files have been uploaded, install the app on an android device (the APK file is stored in document foler), and it will be ready to demo.

# Conclusions & Next steps

This project offers a solution to apply data mining to current digital signage technology. With this structure, data source is not restricted to Facebook data, it could also be Twitter, Instagram or any other user behavior data.

In the future, here are some aspects could be further developed.

1. Data model. Some optimization could be done in the next steps, for example, enlarge training dataset and multi-granularity can be considered. Also if we can have true ground label, there will be more ways to evaluate accuracy, which would help improve solution’s accuracy in the end.
2. Sense schedule logistics. Currently, once a beacon (indicating a screen) detects customers, corresponding screen will be triggered immediately. More complicated scenarios could be considered and sensors are not necessarily ibeacons as well.
3. Substitutes of ibeacons. Based on the pilot performance, we found Bluetooth signal is somewhat weak and unstable. We may consider use another device to play its role if there is a fit.
4. Real-life application. So far this project is still a prototype, some mapping are finished within codes, not via database, for convenience. When it comes to real-life one, a lot work need to be done, and adequate testing are needed too.

# References

1. Customer-centric marketing: https://www.custora.com/university/for-marketers/customer-centric-marketing/basic
2. Amazon Machine Learning: http://docs.aws.amazon.com/machine-learning/latest/dg/
3. Microsoft Azure: http://azure.microsoft.com/en-us/overview/what-is-azure/
4. Predixion Softwate: <https://www.predixionsoftware.com/>
5. Facebook SDK for Android: https://developers.facebook.com/docs/android
6. Facebook Graph API: https://developers.facebook.com/docs/graph-api/using-graph-api/v2.3
7. Facebook API Guidance: https://www.sammyk.me/optimizing-request-queries-to-the-facebook-graph-api
8. Graph API Explorer: <https://developers.facebook.com/tools/explorer/145634995501895/?method=GET&path=me&version=v2.3>
9. Python Regular Expression: https://docs.python.org/2/library/re.html

1. Customer-centric marketing: <https://www.custora.com/university/for-marketers/customer-centric-marketing/basic> [↑](#footnote-ref-1)