Customer Behavioural Analytic on Food Delivery Services

Lee Kang Wenn, Chan Peck Hui and Yang Poh Yee

Abstract—**As penetration of smart phones and mobile applications increases, the technology is going to trigger a massive influx of big data. Consumer analytic is at the epicentre of a big data revolution. Technology helps capture rich and plentiful data on consumer behaviour in real time. Therefore, this project aims to use machine learning to build a more efficient and simpler consumer analytic model for small and medium-sized enterprise that does not have much resources to carry out data analysis.**

Index Terms—**Churn Analysis, Customer behavioural Analysis, Data mining, Decision trees, K-means clustering, Market Segmentation, Neural networks, RFM**

I. Introduction

THE massive increase in the amount of data collected and stored by organizations around the world over the past few decades is evident. In conjunction with this, the ability to access and analyse this data is quickly becoming more and more important. However, some firms do not have the resources to perform consumer analytic and it's often inaccurate [1]. Analysing consumer data often requires huge amount of time, manpower, and professionals, which leads to high cost for small organization to carry out data analysis.

This project aims to achieve several objectives. The main objective of this project is to extract semantic customers behavioural information through data collected from a food delivery company, Running Man. Customer behavioural refers to actions done by customers on the website and number of customers visiting the website.

The data is then processed by going through features extraction, pre-processing and input into machine learning algorithm for output. We will also explore the most accurate prediction model for the project. We will then provide suggestions such as cross-selling products, peak hour prediction and conducting promotions events based on the output from the prediction model.

Data extracted that are trained by machine learning will be examined in terms of predictive accuracy in order to be applied in real-life situation. By examining information such as customers' purchasing behaviour, such information can help the management to make better decision in terms of marketing and planning, hence increases sales performance [2].

Investigating customer behaviour in retail contexts is essential to obtain various formal indicators that are interesting from the marketing research viewpoint such as the conversion rates, to further improve the food delivery experiences. Being able to analyse and predict market and customer behaviour with big data is a new paradigm shift for small and medium-sized enterprises (SMEs). When it is implemented correctly, it can yield increased flexibility, productivity responsiveness, anticipation and ability to meet customer need through capturing blind spots and making better decisions [3].

II. Literature review

More and more companies embrace Artificial Intelligence to assist them in decision making. For example, Ford Motors is using consumer analytic to start its own revolution in product innovation and design [1]. Ford facilitates product innovation in a rapid manner using Big Data without waiting for insights from traditional research such as focus groups and surveys [4].

This solution benefits two parties: the consumer and the merchant. The system can be implemented to provide "personalized shopping experience or services to the consumer". Furthermore, the system can also be implemented to provide sophisticated insights to assist top management in decision making, product perfecting, and more importantly, customer satisfaction. When a firm embraces new technology and constantly seeks for transformation of knowledge, the knowledge creation process grows exponentially. With these resources, a firm can improve performance better than without resources, thus the resources are considered valuable [5].

Before a consumer prediction analytic model is created, an iteration of process is needed.

Figure 1. Process of modelling consumer prediction analytic.

The first step is business understanding. We need to analyse and understand the business logic and the policy of the business in business perspective. Next up is data understanding. We need to study all the available data and investigate the relationship between all available data in order to perform data mining correctly. Data preparation is then carried out to perform ensure all the data is valid for the modelling. In modelling, various techniques are applied into the model and tested. The most suitable technique is then proceed to evaluation. In evaluation, the model is evaluated in different measures such as accuracy, performance and efficiency. Overfitting of model might occur therefore an evaluation of the model is needed to make sure the model can be then be deployed. Lastly, the model will be deployed into real system to start servicing the user.

III. Consumer Behavioural Analytic

Consumer Behavioural Analytic is the analysis of the past data on what the customers do and how they act in a manner that will reflect in what they do and how they will react in the future. Consumer analytic is important to a company, it brings benefits such as understanding customers' need, customers' purchasing power and many more [1]. If there is a strong behavioural analytic exists, it can be widely used in helping to build a smarter businesses with social commerce so that retailers can now record and track the customers on how do they buy, what are their choice criteria, when and how frequent do they buy and their oath channels when checking out products.

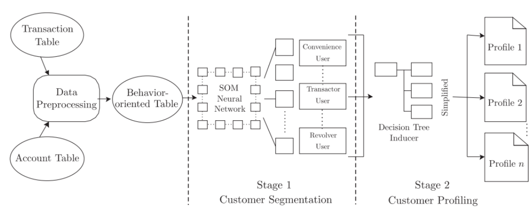

Figure 2. The two-stage framework of consumer behaviour analysis [6].

The study of consumer analytic lies at the junction of Big Data and consumer behaviour. Big data is a hot issue in today's world. 4.4 Zettabytes of data exist in the digital universe today, by 2020, the digital universe is expected to reach 44 zettabytes [7]. Since data provide behavioural insights about consumers; marketers are able to translate those insights into market advantage. Big data is a top business priority and drives enormous opportunity for business improvement [8]. Nevertheless, the first problem is that manually analyses the conglomeration of raw data to gain insights is inefficient and ineffective.

Consumer analytic can be performed by deductive or inductive approaches, where deductive approaches interpret consumer behaviour based on existing theories and model, while inductive approaches do not make any assumptions or hypothesis before the interpretation. Deductive approaches have been widely used, providing good results. However, the need to obtain even more insights has directed marketers' interest towards inductive prediction approaches. Studies have shown that using inductive approaches consumer analytic can advance the understanding of marketing phenomena more compared to using deductive approaches [1]. Without interconnecting the relationship among consumers' purchases, customers' flow and path on the web, deductive approaches would be inaccurate.

IV. Supervised and Unsupervised Machine Learning Algorithm

A. Ordinary Least Square

In statistical modelling, regression analysis is a statistical process which estimates the relationship among variables and the outcome of output for this statistical model is continuous real value, also known as function approximation [9]. There are many regression model that exists, such as Ridge Regression (impose penalty on size of coefficient), Lasso Regression (estimates sparse coefficients).

Ordinary Least Squares, also known as Linear Regression, is one of the simplest regression model in regression models. Linear Regression fits data with the best hyper-plane which "goes through" the points. For each point the differences between the predicted point and the actual observation is the residue.

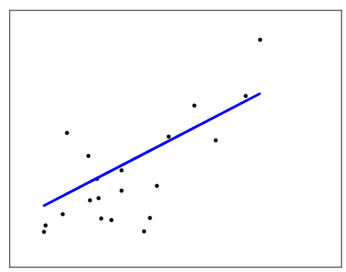

Figure 3. A linear regression visualization.

Linear regression is an approach to find out the sum of error in the dataset. Gradient descent is an iterative method to find the minimum of a function. Gradient descent can also be replaced by gradient ascent to determine the maximum of a function.

Gradient descent starts with an initial set of data; in each step, it decreases each data in proportion to its partial derivative. There will a be gradient descent step size, also known as learning rate, along with features and the data is given as input to the learning algorithm. The partial derivative specifies how much a small change in the data would change the error.

Consider minimizing the sum-of-squares error. The error is a sum over the examples. The partial derivative of a sum is the sum of the partial derivatives. Thus, we can consider each example separately and consider how much it changes the data.

Linear regression computes the least squares solution using a singular value decomposition of X. If X is a matrix of size (n, p) this method has a cost of $O\left(np^2\right)$, assuming that $n \geq p$. If the data used for regression is gigantic, the machine will need longer time to process all the data.

However, Linear regression only limited to linear relationship, which means it only looks at dependent and independent variables. Some of the data might not have straight relationship between two of the variables. Also, linear regression is sensitive to outliers. Outliers will affect the linear relationship plotted thus affect the accuracy of the relationship [10].

Linear Regression is a very powerful statistical technique and can be used to generate insight on the data that we, human usually could not understand without plotting the graph. For example, linear regression can be used in business sector to generate insight on consumer behavior, understanding business and factors influencing profitability. By conducting a linear analysis on the sales data of the company, the company could forecast their sales in the future [11].

B. Support Vector Machine

Support vector machines (SVMs) are a set of supervised learning methods used for classification, regression and outliers detection. It is mostly used in classification problems. In SVM, each data item is plotted as a point in n-dimensional space, where n is the number of features in the dataset) with the value of each feature being the value of a particular coordinate. Then, classification will be performed by finding the hyperplane that differentiate the two classes very well [12].

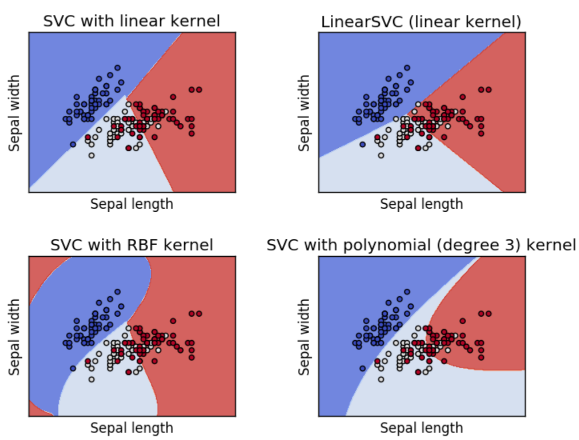

Figure 4. Support Vector Machine visualization.

Support Vectors are simply the coordinates of individual observation. Support Vector Machine is a frontier which best segregates the two classes (hyperplane/line). SVM can identify the right hyperplane by maximizing the distances (margin) between nearest data point (either class).

SVM also has a feature to ignore outliers and find the hyperplane that has maximum margin. Hence, we can say, SVM is robust to outliers.

SVM can also solve problem where the dataset could not have linear hyperplane between the classes. It solves this problem by introducing an additional feature. The machine replot the plane based on the new feature - equation on axis x and z, for example.

C. Random Forest

Decision Trees are a non-parametric supervised learning method used for classification and regression. Decision tree learning is one of the most successful techniques for supervised classification learning. Decision tree model aims to create a model that predicts the value of a target variable by learning simple decision rules inferred from the data features.

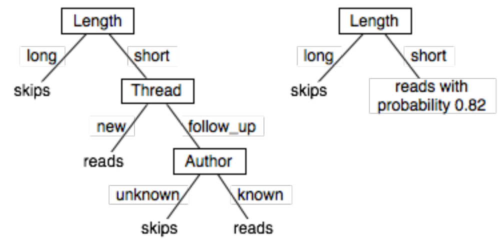

Figure 5. A decision tree visualization.

A decision trees, also known as classification tree, is a flowchart like tree structure, which each internal (non-leaf) node is labeled an input feature. The branch (arcs coming from a node) represents an outcome of the test. Each leaf of the tree is labeled with a class name or class distribution. Random forest contain a forest of decision tree [9].

Random Forest belongs to a larger class of machine learning algorithms called ensemble methods. Ensemble learning involves the combination of several models to solve a single prediction problem.

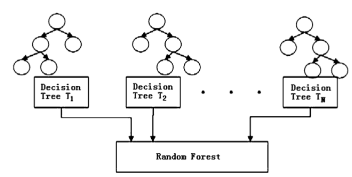

Figure 6. A random forest visualization.

In random forests, each tree in the ensemble is built from a sample drawn with replacement (eg. bootstrap sample) from the training set. In addition, when splitting a node during the construction of the tree, the split that is chosen is no longer the best split among all features. Instead, the split that is picked is the best split among a random subset of the features. As a result of this randomness, the bias of the forest usually slightly increases (with respect to the bias of a single non-random tree) but, due to averaging, its variance also decreases, usually more than compensating for the increase in bias, hence yielding an overall better model.

D. Artificial Neural Network

An Artificial Neural Network (ANN) is an information processing paradigm that is inspired by the way biological nervous systems, such as the brain, process information [13]. The key element of this paradigm is the novel structure of the information processing system. It is composed of a large number of highly interconnected processing elements (neurons) working in unison to solve specific problems. ANNs, like people, learn by example. An ANN is configured for a specific application, such as pattern recognition or data classification, through a learning process. Learning in biological systems involves adjustments to the synaptic connections that exist between the neurons. This is true of ANNs as well.

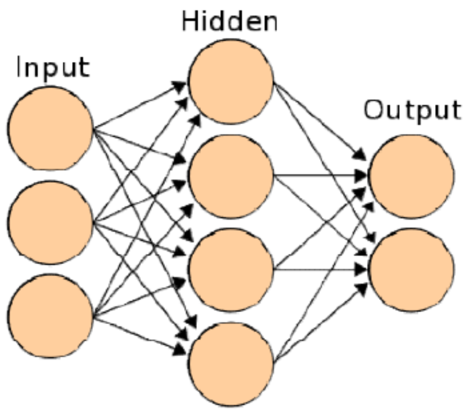

Figure 7. Artificial Neural Network visualization.

The input is the number of input, the data inputted will be push into hidden layer, and the hidden layer will process the information and output the data. In neural network, the input layer, hidden layer and output layer are similar as our human brain, where nerves is nodes in this case.

E. K-means clustering

K-means [14] is one of the simplest unsupervised learning algorithms that solve the well known clustering problem. The procedure follows a simple and easy way to classify a given data set through a certain number of clusters (assume k clusters) fixed a priori. The main idea is to define k centroids, one for each cluster.

It is typically used for scenarios like understanding the population demographics, market segmentation, social media trends, anomaly detection, etc. where the clusters are unknown to begin with.

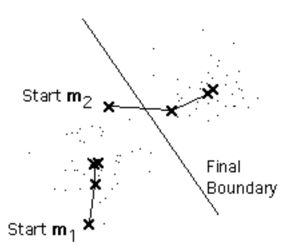

Figure 8. Analytic Process Model visualization.

In training phase of K-Means, K observations are arbitrarily selected (known as centroids). Each point in the vector space is assigned to a cluster represented by nearest (euclidean distance) centroid. Once the clusters are formed, for each cluster the centroid is updated to the mean of all cluster members. And the cluster formation restarts with new centroids. This repeats until the centroids themselves become mean of clusters, i.e., when updating centroids to mean doesn't change them. The prediction of a test observation is done based on nearest centroid.

V. Research Methodology

A. Analytic Process Model

Prediction model is part of analytics process in data science. It is a model which study historical data to forecasting and make prediction for future. The prediction model is using statistics number to show the likelihood of that particular prediction will happen in future. Table 2.3.1 below with diagram and description will describe the whole process from data collection stage until prediction making stage.

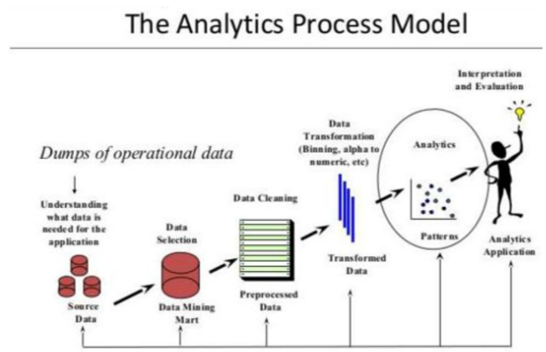

Figure 9. Analytic Process Model visualization.

Table I
The Analytic Process Model and explanation.

Process	Explanation
Source Data	Unstructured data from various places such as: excel files, log book
Data Mining	Retrive the data from source data and discover their correlation and pattern.
Preprocess Data	Data filtering process to ensure data is having minimal error before proceed to the next step. For example, deciding what to do with missing data - discard outlier or replace with values.
Transform Data	Because data is stored in various form and format, therefore data transformation is needed to transform the data into suitable data format for machine learning.
Analytic	Visualizing the information so that it's easier to understand, in terms of diagrams such as histogram, bar chart.
Analytic Application	Extract knowledge from the visualized information and apply the knowledge into application.

B. The dataset

The dataset that we managed to collect are the transaction data from a food delivery services. Here are the columns available in the data set.

C. Predictive Analysis

1) Recency Frequency Monetary: RFM (recency, frequency, monetary) analysis is a marketing technique used to

Table II
DATA DESCRIPTION

Column Name	Data type	Description
id	String	User's id
address.city	String	User's city
address.street	String	User's location address
comments	String	User's comment when checking out E.g. No seasonings added in the porridge / Please prepare small change
contact.email	String	User's email
contact.fname	String	User's first name
contact.lname	String	User's last name
contact.number	String	User's contact number
coordinate.lat	Double	User's location in coordinate latitude
coordinate.lng	Double	User's location in coordinate longitude
delivery time.date	Date	User's delivery date
delivery time.schedule	Time	When to deliver the food (ASAP or later)
delivery time.time	Time	User's order delivery time
payment.amount	Double	User's order amount
transaction date	Date	User's order date
transaction id	String	User's order id
task	String	Items to be delivered, to send to another part of the system, including the details such as shop name, shop coordinate, task deliver to and etc.

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). RFM analysis is based on the marketing axiom that ”80% of your business comes from 20% of your customers.”

For more than 30 years, direct mailing marketers for non-profit organizations have used an informal RFM analysis to target their mailings to customers most likely to make donations. The reasoning behind RFM was simple: people who donated once were more likely to donate again. With the advent of e-mail marketing campaigns and customer relationship management software, RFM ratings have become an important tool. Using RFM analysis, customers are assigned a ranking number of 1,2,3,4, or 5 (with 5 being highest) for each RFM parameter. The three scores together are referred to as an RFM ”cell” . The database is sorted to determine which customers were ”the best customers” in the past, with a cell ranking of ”555” being ideal.

Although RFM analysis is a useful tool, it does have its limitations. A company must be careful not to oversolicit customers with the highest rankings. Experts also caution marketers to remember that customers with low cell rankings should not be neglected, but instead should be cultivated to become better customers.

2) Market Segmentation: Market Segmentation is a process of dividing the customers into different groups and segments on the basis of certain characteristics [15]. Common characteristic such as city they stayed, type of food they ordered and period of time they made an order will be looked into in segmenting the customers. The reason of market segmentation is to help the company in creating a marketing mix strategy for each segment and cater them accordingly. By using the STP (Segmentation, Targeting and Positioning) strategy to divide the marketplace into different segments that require different marketing mixes, then review the market segments and deciding which to pursue and position their offer more effectively for the customers.

3) Churn Analysis: Churn analysis can be defined as a predictive techniques to analyse the percentage of subscribers to a service who discontinue their subscriptions to that service within a given time period [16]. Many firms used Churn analysis to assess their customers' value in order to retain or even cultivate the profit potential of customers [17]. Figure shows the work-flow of Churn Prediction Analysis.

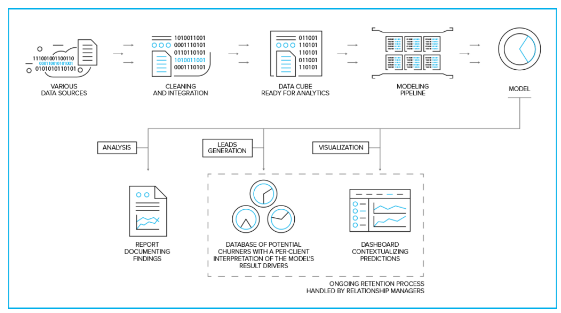

Figure 10. Churn Prediction Analysis.

ACKNOWLEDGEMENT

The completion of this study could not have been possible without the expertise of Dr Chaw Jun Kit, our beloved supervisor and Professor Lim Tong Ming for encouragement and guiding us during the entire project development.

A debt of gratitude is also owed to the company that provide the customer transaction dataset for us to process in this project.

Next, we also appreciate TAR University College that giving us support in lab facilities to conduct our project and transportation to participate the training provided by dreamcatcher.

Last but not the least, we are grateful to have each other as team mates and always having discussion and meetings for problems solving and work contribution. Thanks to the team cooperation,we had successfully developed the project.

References

[1] S. Erevelles, N. Fukawa, and L. Swayne, “Big data consumer analytics and the transformation of marketing,” *J. Bus. Res.*, vol. 69, no. 2, pp. 897–904, 2016. [Online]. Available: http://dx.doi.org/10.1016/j.jbusres.2015.07.001

[2] A. Raorane and R. V. Kulkarni, “Data mining techniques: A source for consumer behavior analysis,” Sep. 2011. [Online]. Available: http://arxiv.org/abs/1109.1202

[3] D. Sen, M. Ozturk, and O. Vayvay, “An overview of big data for growth in SMEs,” *Procedia - Social and Behavioral Sciences*, 2016. [Online]. Available: http://dx.doi.org/10.1016/j.sbspro.2016.11.011

[4] G. Satell, “5 things managers should know about the big data economy,” https://www.forbes.com/sites/gregsatell/2014/01/26/5-things-managers-should-know-about-the-big-data-economy/, Jan. 2014, accessed: 2017-11-15. [Online]. Available: https://www.forbes.com/sites/gregsatell/2014/01/26/5-things-managers-should-know-about-the-big-data-economy/

[5] I. V. Kozlenkova, S. A. Samaha, and R. W. Palmatier, “Resource-based theory in marketing,” *J. of the Acad. Mark. Sci.*, vol. 42, no. 1, pp. 1–21, Jan. 2014. [Online]. Available: http://link.springer.com/10.1007/s11747-013-0336-7

[6] N. C. Hsieh and K. C. Chu, “Enhancing consumer behavior analysis by data mining techniques,” *Int. J. Inf. Manage. Sci.*, 2009. [Online]. Available: http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.579.9503&rep=rep1&type=pdf

[7] IDC, “The digital universe of opportunities: Rich data & the increasing value of the internet of things. EMC2,” https://www.emc.com/collateral/analyst-reports/idc-digital-universe-2014.pdf, 2014, accessed: 2017-11-13. [Online]. Available: https://www.emc.com/collateral/analyst-reports/idc-digital-universe-2014.pdf

[8] b. J. Kennedy, “Amount of data in 2011 equal to 57.5bn 32GB apple ipads - enterprise — siliconrepublic.com - ireland’s technology news service,” https://www.siliconrepublic.com/enterprise/amount-of-data-in-2011-equal-to-57-5bn-32gb-apple-ipads, Jun. 2011, accessed: 2017-11-13. [Online]. Available: https://www.siliconrepublic.com/enterprise/amount-of-data-in-2011-equal-to-57-5bn-32gb-apple-ipads

[9] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and É. Duchesnay, “Scikit-learn: Machine learning in Python,” *J. Mach. Learn. Res.*, vol. 12, no. Oct, pp. 2825–2830, 2011. [Online]. Available: http://www.jmlr.org/papers/volume12/pedregosa11a/pedregosa11a.pdf

[10] Sciencing, “The disadvantages of linear regression,” http://sciencing.com/disadvantages-linear-regression-8562780.html, 2017, accessed: 2018-1-13. [Online]. Available: http://sciencing.com/disadvantages-linear-regression-8562780.html

[11] Z. Xu, G. L. Frankwick, and E. Ramirez, “Effects of big data analytics and traditional marketing analytics on new product success: A knowledge fusion perspective,” *J. Bus. Res.*, 2016. [Online]. Available: http://dx.doi.org/10.1016/j.jbusres.2015.10.017

[12] S. Ray, P. Kaur, A. Choudhary, F. Shaikh, and NSS, “Understanding support vector machine algorithm from examples (along with code),” https://www.analyticsvidhya.com/blog/2015/10/understaing-support-vector-machine-example-code, Sep. 2017, accessed: 2018-1-13. [Online]. Available: https://www.analyticsvidhya.com/blog/2015/10/understaing-support-vector-machine-example-code

[13] D. F. Specht, “Probabilistic neural networks,” *Neural Netw.*, vol. 3, no. 1, pp. 109–118, Jan. 1990. [Online]. Available: http://www.sciencedirect.com/science/article/pii/089360809090049Q

[14] J. MacQueen and Others, “Some methods for classification and analysis of multivariate observations,” in *Proceedings of the fifth Berkeley symposium on mathematical statistics and probability*, vol. 1, 1967, pp. 281–297. [Online]. Available: https://books.google.com/books?hl=en&lr=&id=IC4Ku_7dBFUC&oi=fnd&pg=PA281&dq=MacQueen+1967&ots=nOUeK1JfrK&sig=g1UPl5pyEN7yB5Ofen1VT1zOEzc

[15] A. Dutta, “Market segmentation - definition, basis, types & examples — feedough,” https://www.feedough.com/market-segmentation-definition-basis-types-examples/, Feb. 2018, accessed: 2018-4-16. [Online]. Available: https://www.feedough.com/market-segmentation-definition-basis-types-examples/

[16] C.-F. Tsai and M.-Y. Chen, “Variable selection by association rules for customer churn prediction of multimedia on demand,” *Expert Syst. Appl.*, vol. 37, no. 3, pp. 2006–2015, 2010. [Online]. Available: http://dx.doi.org/10.1016/j.eswa.2009.06.076

[17] M.-K. Kim, M.-C. Park, and D.-H. Jeong, “The effects of customer satisfaction and switching barrier on customer loyalty in korean mobile telecommunication services,” *Telecomm. Policy*, vol. 28, no. 2, pp. 145–159, 2004. [Online]. Available: http://dx.doi.org/10.1016/j.telpol.2003.12.003