Journal of Consumer Sciences **E-ISSN: 2460-8963**
2021, Vol. 06, No. 01, 1-19

Analysis of Demography, Psychograph and Behavioral Aspects of Telecom Customers Using Predictive Analytics to Increase Voice Package Sales

Billy Goenandar*)[1], Maya Ariyanti[2]
[1,2]Magister Management, Faculty of Economy and Business, Telkom University, Bandung 40152, West Java, Indonesia

*Corresponding author: goenandar.billy@gmail.com

Abstract

In 2018, Telkomsel's core business shifted its main services from Telephone and SMS services to Data and Digital services, since a declining trend of revenue starting 2014. However, telephone service still contributed 28.4% to the revenue and was the second largest, while SMS gave 4.1%. This research predicts voice package buyers using predictive analytics to identify customer profiles and significant variables to form appropriate target customer segmentation. Logistic regression was used to predict customers who would buy voice packages using 15 input variables. Next, analytics was done by dividing the data into 70% training data sets and 30% testing data obtained from customer voice package user data. The model accuracy gained 97.2%, and the top seven significant variables were formed. Then five clusters of customer segmentation were formed based on top significant variables using the K-Means clustering technique. Based on the results of the prediction model and clustering, behavioral targeting was conducted to provide targeted gimmick products based on five segmentations formed, and then it was divided into two main target customers by considering the similarity of behaviors based on revenue voice, minutes of voice usage, voice transactions, day of voice usage and data payload, thus it was more targeted.

Keywords: behavioral targeting, personalized marketing, predictive analytics, segmentation

Abstrak

Pada tahun 2018, Core business Telkomsel berubah dari layanan Telpon dan SMS menjadi layanan Data dan Digital karena tren revenue yang terus menurun sejak 2014. Namun kontribusi revenue layanan telepon masih di angka 28.4% dan SMS 4.1%. Penelitian ini memprediksi pembeli *voice package* dengan menggunakan *predictive analytics* untuk mengidentifikasi profil pelanggan dan *significant variable* untuk menentukan target pelanggan yang sesuai. *Logistic regression* digunakan untuk memprediksi pelanggan yang akan membeli paket *voice package* dengan menggunakan 15 variabel input. Setelah itu *analitycs based table* dilakukan dengan membagi data set menjadi 70% *training data Set* dan 30% *testing data set* dari data pelanggan pengguna *voice package*. Akurasi model didapatkan 97.2% dan membentuk 7 variabel signifikan teratas. Kemudian lima cluster segmentasi pelanggan berdasarkan terbentuk menggunakan teknik *clustering K-Means*. Berdasarkan hasil dari *prediction model* dan

© 2021 The Author(s). This open access article is distributed under a Creative Commons Attribution (CC-BY) 4.0 license

clustering tersebut, *behavioral targeting* dilakukan untuk memberikan *targeted gimmick product* berdasarkan lima segment yang dibagi menjadi 2 target pelanggan utama dengan mempertimbangkan kemiripan perilaku berdasarkan *revenue voice, minutes of usage voice, transaksi voice, day of usage voice* dan *data payload* sehingga lebih tepat sasaran.

Kata kunci: analitik prediktif, pemasaran personal, sasaran perlakuan, segmentasi

Introduction

As the 6th largest cellular operator in the world in terms of the number of subscribers, Telkomsel is the market leader in Indonesia's telecommunications industry, which was now trusted to serve 163 million subscribers by December 2018. Based on Telkomsel's 2018 annual report data (Telkomsel, 2019), Telkomsel recorded a profit of Rp 25.5 trillion and a net income margin of 28.6%. However, in the same year, Telkomsel recorded a 4.3% decrease in revenue to Rp 89.2 trillion due to the accelerated transition of Legacy to data services and the strict competition among cellular operators in Indonesia.

Changes in consumer behavior certainly affect the services provided to the customers and the revenue generated. Core business changes from Telephone and SMS services to Data and Digital services. Telephone and SMS have become legacy businesses with a declining trend in revenue, starting in 2014. However, the contribution of telephone service revenue was still 28.4%, while SMS was still at 4.1%, as presented in the following chart:

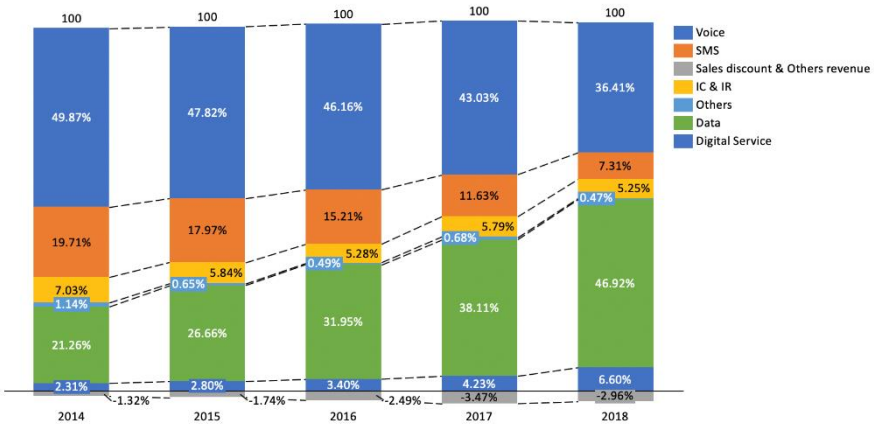

Figure 1 Proportion of revenue contribution (Source: Annual report 2014-2018)

Based on Figure 1, the behavior change has made changes to the portion of revenue in Telkomsel. In 2014 the proportion of revenue from voice was still dominant at 49.8%, along with industrial changes. Four years later, the proportion changed to 36.41% or was the second largest. It was inversely proportional to revenue data. In 2014, the proportion of revenue data was 21.26%. Four years later, the proportion of revenue was the largest at 46.92%. The decline in Legacy has also been accelerated by the growing replacement of telephone services, namely over the top (OTT) delivery (Franendya, 2019).

With the continued decline in telephone service revenue, it has been predicted that the decline will occur in the fourth quarter of 2019 by -5.4% compared to the third quarter of 2019. This is due to the greater decreasing trend in the fourth quarter of each year.

There are two types of telephone services, namely PAYU (pay as you use) and voice package services. PAYU service is a payment system to customers based on a certain tariff system that will be charged according to the duration used. While invoice packages, customers buy the bulk duration of certain phone calls according to a predetermined scheme. The two services' performance shows the difference that is the voice package had better performance than PAYU. This is illustrated in the performance of telephone services in January-September 2019 in the following chart.

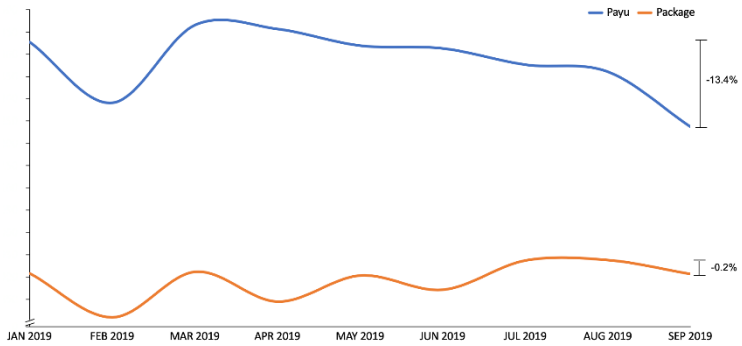

Figure 2 Performance of 2nd layer telephone service (Source: Telkomsel Internal)

Figure 2 illustrates that the decline in PAYU was greater than that of the voice package by 13.4% in September compared to January 2019. In comparison, the voice package experienced a decrease, which tended to be better by 0.2% in September compared to January 2019. This is supported by the proportion of the 2nd layer telephone service revenue. PAYU service continues to erode along with the decline that continued while the package service portion increased. However, this was not enough to cover the decline in PAYU revenue, as illustrated in Figure 3.

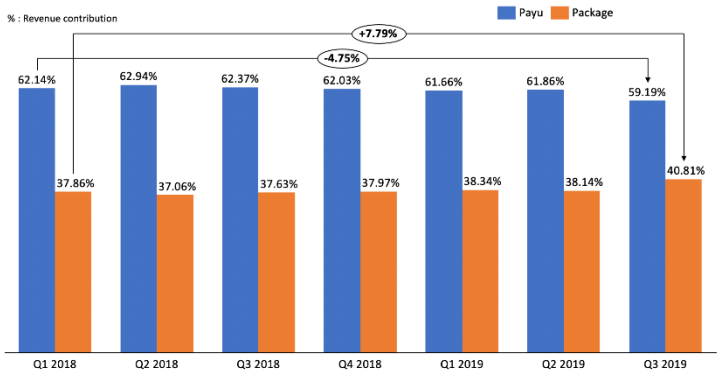

Figure 3 Contribution of 2nd layer telephone service revenue (Source: Telkomsel Internal)

Nevertheless, the number of active users of Telkomsel's voice package services was still smaller than that of PAYU or users of PAYU and voice packages. Figure 4 presents the profiles of Telkomsel telephone service users.

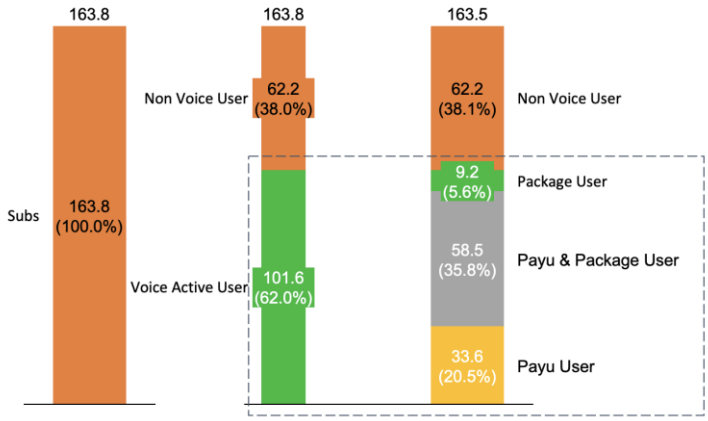

Figure 4 Profile of telephone service users September 2019 (Source: Telkomsel Internal)

As seen in the customer profile, PAYU customers in numbers were still more than the voice package users, and the revenue decrease is larger. From the data above, Telkomsel needs to prepare a special marketing strategy to increase voice package users because voice package revenue performance seems more stable when compared to revenue from PAYU users. This is aimed to suspend the rate of decline in revenue from Telkomsel's telephone service, considering that the contribution of telephone service revenue is still the 2nd largest after data/broadband services, as illustrated in Figure 1.

Literature Review

According to Kotler and Keller (2016), marketing is a process in which a company creates value for its customers and builds a stronger relationship with the customers to gain feedback. Some researchers have even questioned the main assumption of personalized marketing, stating that marketers can reveal customers' preferences to build a learning relationship (Ma et al., 2017; Ebrahim et al., 2016). Amidst the business competition, a company must understand customer needs to provide extra value for the customers. The more customers obtain, the higher their transactions, thus contributing to the company's revenue. The purpose of marketing is to draw new customers, provide more value or benefit for the customers, and maintain existing customers by keeping them satisfied.

In the development of digital technology, marketing also evolves into digital marketing. Chafey (2013) stated that digital marketing is a way to achieve marketing goals by applying digital technology, from identifying the customers, recognizing their needs, building products with higher values perceived by the customers, and conducting promotional activities to the customers. The adaptation of digital technology into marketing can make marketing more effective, efficient, and wide-ranged.

Marketing strategy involves two main questions. The first question is, "Which customers will be served (segmentation and targeting)?" The second question is, "How will the company create value for them (differentiation and positioning)?" The company then designs a marketing program by providing the value desired by targeted customers (Kotler & Keller, 2016). The concept of marketing is a business philosophy that states that satisfying customer needs is an economic and social condition for the company's survival (Swastha & Irawan, 2005). Meeting the needs and desires of the customers requires a marketing concept, namely the Marketing mix. According to Kotler & Keller

(2016), a marketing mix is a combination of four important variables from the marketing concept that the company can control. The four variables are product, price, place and promotion, and commonly abbreviated as 4P.

According to Goldsmith (1999), strategy and marketing management theory needs to develop and change according to changing markets and marketing practices. The marketing mix changed to 7Ps by adding personnel, physical assets & procedures (Goldsmith, 1999). Nowadays, marketing is more focused on building relationships with customers, called personalization. The essence of this activity is to provide goods and services according to individual needs. Personalization becomes important in marketing strategy and management to become an additional element in the marketing mix besides the product, price, place, promotion, personnel, physical asset, and procedures to become the new 8Ps marketing mix (Goldsmith, 1999). Furthermore, personalization occurs when the firm decides what marketing mix is suitable for the individual—usually based on previously collected customer data (de Bellis et al., 2019).

Concerning marketing strategy, behavioral targeting customizes messages to individual consumers based on their specific shopping interests and characteristics like gender, age, and ethnicity (Ozcelik & Varnali, 2018). Behavioral targeting links identifiers within consumers' browsers to track browsing behavior. This digital tag does not identify a consumer by name. However, it functions more like an index connecting non-contiguous browsing sessions.

According to Vesanen (2005), personalized marketing uses technology and customer information to individually adjust electronic commerce interactions between businesses and customers. Furthermore, with customers of ever-increasing technology savvy, personalized products and services become a business necessity (Shen, 2014). Using the information previously obtained in real-time about the customer profile is then used to offer products or services according to customer needs. Similarly, personalization simply means individualizing customers' shopping experiences based on data collected about them by marketers (Dangi & Malik, 2017). Thus, one of the keys to success in carrying out personalized marketing is technology and customer data. With the development of information technology that can quickly collect customer data and process data, the customer profile from both descriptive and predictive analytics will be valuable in offering targeted products according to customer needs. Here is an overview of the personalized marketing process:

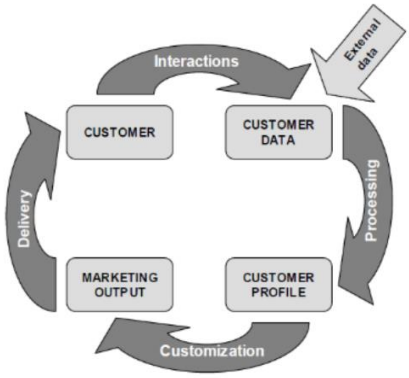

Figure 1. Personalized Marketing Process (Vesanen, 2005)

Methods

According to the above elaboration of theoretical foundation, this research seeks to adapt Goldsmith's 8Ps (1999). The personalized marketing used in this research refers to Vesanen (2005), stating that using technology and information is to adapt the interaction of electronic trade between business and customer as individuals. In other words, this study seeks to predict voice package buyers using predictive analytics to identify customer profiles and significant variables to form appropriate target customer segmentation.

Participants

In collecting data, the author used Telkomsel data related to customer behavior such as conversation data, SMS, top up, purchase of voice packages using an average of three months of data for July-September 2019. This research employs a quantitative descriptive research design to conduct productive analytics using logistic regression. This method aims to predict Voice Payu Users who will buy Voice Packages or who will not buy voice packages by creating a model based on Telkomsel customer transaction data from a population of 0.6 million voice users in September 2019 Jabotabek (Jakarta, Bogor, Tangerang, and Bekasi) region. The data were obtained using the primary data collection method. Afterward, the analytical based table is made using a random sampling method. Furthermore, customer segmentation was conducted using the K-Means method, which later provided voice package offers that follow each segment's characteristics.

This research used primary collection methods because the data used were internal data of the research object and did not include external data. This research used all Telkomsel prepaid subscribers, as many as 0.6 million, who had become consistent Voice RGB users for three months in the central Jabotabek region, where the regional performance is the worst among other regions. The sample was taken from the population using random sampling. Of all the subscribers, 0.4 million were Voice Payu Only users who could buy Voice Package, and 0.2 million were Voice Package subscribers. For the analytical base, the sample was divided into two data sets, namely, training data set with a proportion of 70% and testing data set with a proportion of 30%.

Measurement

In analyzing data requirements, the required data sources described as follows. This research chose the variables used in Fudin (2018) and Kurniawan (2019) research on predictive analysis. Their research focuses on mobile data, while this research emphasizes the use of voice packages. Therefore, the following variables are selected:

1. MSISDN (key variable)
2. Predictable variable, namely the status of active voice PAYU users who bought voice packages and did not buy voice packages (Target variable-Y)
3. Telkomsel subscriber demographic variable (input variable-X)
4. Telkomsel subscriber psychographic variable (input variable-X)
5. Telkomsel subscriber behavioral variable that is related to billing, data calls, voice package behavior and behavior recharge, namely the number of credit purchases by subscribers (input variable-X).

In practice, MSISDN, which is a key variable, did not affect customer behavior, while the predictable (target) variable used was the status of purchasing voice package and not purchasing voice package. Thus, the variables required were defined in the form

of a data mart. The data mart was later formed through an aggregation process in line with the predictive modeling's needs using the Hadoop Data Warehouse tools in Telkomsel. The variables used in this research were from the period of January-March 2020 as many as 25 variables.

Analysis

These data went through an aggregation process following the need to create predictive modeling. The aggregation process and preparation of the analytics data mart used the Hadoop data warehouse tool.

In preparation of analytics-based table as baseline to build predictive modeling, data cleansing was performed and gained the number of clean data of 0.5 million subscribers out of 0.6 million and the number of non-data package purchaser of 0.4 million. In this study, the sample was Payu voice customers only in the central Jabotabek area of 0.4 million, which was included in buying voice packages and at a rate of 0.2 million customers who bought voice packages. The authors made an analytical-based table using the random sampling method from 0.6 million subscribers & divided it into two data sets: (1) Training data set with a proportion of 70%; and (2) Testing the data set with a proportion of 30%.

1. Iteration 1

70% random samples of non-voice package purchasers as many as 71,147 subscribers and 30% random samples of voice package purchasers as many as 166,010 subscribers were then assigned to training and testing data sets respectively.

Table 1 Analytics based on iteration 1

Status Voice Package	Training Data Set	Testing Data Set	Total
	70%	30%	
0	49.763	21.384	71.147
1	116.100	49.910	166.010
Total	158.853	71.294	237.157

2. Iteration 2

50% random samples of non-voice package purchasers as many as 163,689 subscribers and 50% random samples of voice package purchasers as many as 166,010 subscribers were then assigned to training and testing data sets respectively.

Table 2 Analytics based on iteration 2

Status Voice Package	Training Data Set	Testing Data Set	Total
	70%	30%	
0	114.524	49.165	163.689
1	116.100	49.910	166.010
Total	158.853	71.294	329.699

The research began with the stages of business understanding, data understanding, and data preparation. At the modeling stage, it was continued by making a predictive model using the logistic regression method and creating clustering using the K-MEANS method, one of the methods in the IBM-SPSS Modeler version 18 software. The data has collected from Telkomsel data related to customer behavior such as conversation data,

SMS, top up, purchase of telephone packages using an average of 3 months of data from July to September 2019. These data have gone through an aggregation process following the need to create predictive modeling. The process of aggregation and preparation of analytics data marts uses the Hadoop data warehouse tools

1. Predictive Analytics

According to Lin, Ke, & Tsai (2017), 33 techniques in data mining have been identified from these 8 business areas, as follows: bankruptcy prediction, customer relationship management, fraud detection, intrusion detection, recommender system, software development, stock prediction, others financial area, and most of its problem is on prediction. From several existing statistical modeling techniques, the logistic regression was chosen to make prediction. According to Kotu & Deshpande (2015), logistic regression predicts the target variable's value in binary (0 or 1) by using a numeric type input variable. In this research, the target variable (Y) is the status of active voice PAYU subscribers who bought voice packages and did not buy voice packages. The input variable (X) in this research is the demographic, psychographic, and language variables of Telkomsel subscribers.

Hosmer et al. (2013) explain that the regression method has become an integral component of any data analysis related to describing the relationship between the response variable and one or more explanatory variables.

The output of the analysis using the logistic regression method was to predict active voice PAYU users who have high potential to purchase Telkomsel voice packages with purchasing scores.

2. Logistic Regression

From several existing statistical modeling techniques, this research uses logistic regression. According to Kotu & Deshpande (2015), logistic regression predicts the target variable's value in binary (0 or 1) using a numeric type input variable. In this research, the target variable (Y) is the status of active voice PAYU users who bought voice package and did not buy voice package. The input variable (X) is the demographic, psychographic, and behavior variables of Telkomsel subscribers.

In analyzing the research findings, the validity measure was conducted on the model made using the logistic regression. The confusion matrix as in Table 3 was used to measure the model's validity whose target value was binary (0 or 1) such as in logistic regression (Wendler, & Gröttrup, 2016). The values in the confusion matrix were obtained from the results of model testing.

Table 3 Confusion Matrix

	Actual Condition		
	Total Population	**Purchaser**	**Non-Purchaser**
Prediction	Purchaser	TP (True Positive)	FP (False Positive)
	Non-Purchaser	FN (False Negative)	TN (True Negative)

Source: (Wendler, & Gröttrup, 2016)

True positive rate (TPR) or sensitivity = $\frac{TP}{TP+FN}$

True negative rate (TNR) or spesificity = $\frac{TN}{FP+TN}$

$Accuracy$ (ACC) $= \frac{TP + TN}{Total\ Populasi}$

The higher the model's accuracy and sensitivity values , the better the model is to be selected. As for the validity of clustering using K-Means, silhouette index was used. The silhouette index measures each object's average distance with other objects in a cluster, including measuring the average distance from one cluster to another (Wendler, & Gröttrup, 2016).

3. Clustering

After the score was calculated using logistic regression, the segmentation of customers who are predicted to purchase voice packages was made using clustering namely K-Means in SPSS Modeler version 18. K-Means is the most frequently used clustering method to determine the number of clusters (k) of a data set. K-means determines the center point of each cluster called the centroid. K-means clustering creates K partitions in n-dimensional space, where n is the number of attributes in a data set. To partition a data set, it is necessary to define a measure of proximity. The most commonly used measure for attributes is the Euclidean distance (Kotu & Deshpande, 2015).

One of the advantages of the K-Means clustering method is that although allocating absolute cluster membership in the data is possible, this can be done at a better level of granularity by providing a membership percentage (Pal & Bhattacherjee, 2015: 65).

The similarity in subscribers' profile who would purchase voice packages in one cluster can be found and significantly differentiated from other clusters by creating clustering. The stages in making clustering with the K-Means method based on the results of logistic regression were:

i. Identifying all input variables that will be used in clustering. In this research, there are 25 input variables or predictors;
ii. Determining the significant variable based on the results of logistic regression;
iii. Determining the K value of K-Means, namely the number of clusters that will be made for each significant variable; and
iv. Validating cluster accuracy by measuring the silhouette index.

The input variables used in the clustering model formation were the top seven significant variables from the logistic regression results. Furthermore, for the validity of the clustering results from K-Means, a silhouette index was used to measure each object's average distance from other objects in one cluster, including measuring the average distance from one cluster to another (Wendler, & Gröttrup, 2016). SPSS Modeler software version 18 showed the silhouette index of the cluster of 0.4 and is in the fair category.

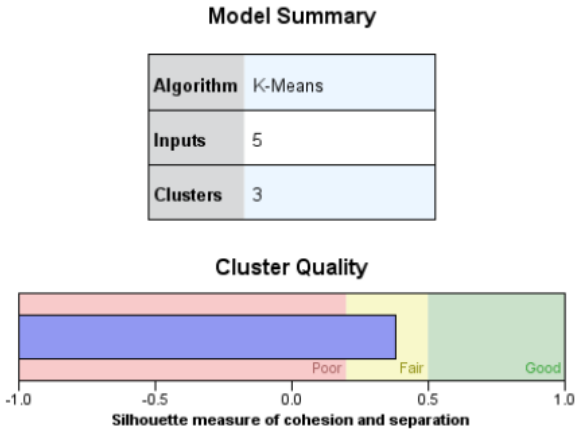

Figure 5 The silhouette index

The silhouette index measures each object's average distance with other objects in a cluster, including measuring the average distance from one cluster to another (Wendler, & Gröttrup, 2016). The similarity of customer profiles in one cluster can be found and significantly differentiated from other clusters by clustering.

Findings

Predictive Analytics

Iteration 1

Table 4 Confusion Matrix Iteration 1

Actual Condition				Actual Condition			
	Total Population	Purchaser	Non-Purchaser		**Total Population**	Purchaser	Non-Purchaser
Prediction	Purchaser	113,844	2,256	**Prediction**	Purchaser	48,938	972
	Non-purchaser	2,781	46,982		Non-purchaser	1,176	20,208
Accuracy	97.0%			Accuracy	97.0%		
Sensitivity	97.8%			Sensitivity	97.8%		

Table 4 above presents that the accuracy of the logistic regression model reached 97.0%. In addition, the measurement of logistic regression validity was done using confusion matrix, which measures the accuracy of the model and the sensitivity of the model.

Table 5 Significant variable iteration 1

No	Variable Input	Normalized Sensitivity Score
1	Voice_rev	0.46
2	Mou_voice	0.23
3	Trx_voice	0.18
4	Payload_data	0.06
5	Sms_rev	0.02
6	Trx_sms	0.02
7	Dou_voice	0.01

Iteration 2

Table 6 Confusion matrix iteration 2

	Actual Condition				Actual Condition		
	Total Population	Purchaser	Non-Purchaser		Total Population	Purchaser	Non-Purchaser
Prediction	Purchaser	111,676	4,424	Prediction	Purchaser	47,978	1.932
	Non-purchaser	2,131	112,517		Non-purchaser	856	48,362
Accuracy	97.2%			Accuracy	97.2%		
Sensitivity	98.2%			Sensitiviy	98.2%		

Table 7 Significant variable iteration 2

No	Variable Input	Normalized Sensitivity Score
1	Voice_rev	0.42
2	Mou_voice	0.24
3	Trx_voice	0.18
4	Payload_data	0.06
5	Sms_rev	0.04
6	Dou_voice	0.01
7	Type_handset	0.01

Based on the confusion matrix for the iteration, iteration 2 had the highest accuracy and sensitivity score and applied it to all populations of non-voice package purchasers of 349.006 customers. There were identified 144.342 customers as highly likely to purchase voice packages based on predictive modeling using logistic regression.

Customer Segmentation

The customer segmentation was developed based on 144.342 customers who had a high likelihood of purchasing a voice package using the clustering method (K-Means). K-Means is one of the clustering algorithms aim at dividing data into groups, hence, the grouped data are based on its characteristics, which are categorized into same cluster, and the data with different characteristics are in different groups (Widiyaningtyas et al., 2017). K-Means is the most commonly used clustering method to determine the number of clusters (k) of a set of data. Five clusters were identified.

Based on Table 8, 60.0% of customers who were identified as having high likelihood to purchase voice package came from Low-Early Internet Adopter segment with average voice revenue per user of IDR 9,017 and day of voice about 4 days.

At present the term "one size fits all" is no longer valid to streamline marketing strategies, as described in the results of research Hjort, Lantz, Ericsson & Gattoma (2013). Using the customer segmentation, companies can carry out various marketing strategies based on related consumer behavior to avoid excessive costs and sent targeted messages and offers to customers. In addition, according to Tuckwell (2016), there are four levels of market segmentation, one of which is Direct or One to One segmentation which is designed according to customer needs and preferences.

Table 8 Customer segmentation

No	Segment Name	Cluster		Revenue				Other Parameters				Handset Type
		Size	Size %	All Service (IDR)	Data (IDR)	Voice (IDR)	SMS (IDR)	MoU Voice (Min)	TRX Voice (time)	Dou Voice (Day)	Payload Data (Gb)	
1	Low-early Internet adopter users	86,535	60.0%	40,615	29,847	9,017	1,751	11	6	4	2.0	Smartphone
2	High-traditional users	6,922	4.8%	73,346	2,551	69,342	1,452	84	39	16	0.1	Non Smartphone
3	Internet user	13,221	9.2%	200,239	188,848	9,371	2,020	10	7	4	27.1	Smartphone
4	Mid-traditional user	20,805	14.4%	22,026	3,124	16,648	2,254	22	10	6	0.2	Non Smartphone
5	Mid-early Internet adopter user	16,859	11.7%	112,940	50,967	59,892	2,081	73	33	14	4.0	Smartphone

Behavior targeting

In this research, mapping of target customers were done based on five pre-formed segmentations by considering similarities. The mapping is presented as follows.

Table 9 Customer targeting

No	Segment Name	Cluster Segmentation	Other Parameters					Handset Type	Target Group
		Size %	Rev Voice (IDR)	MoU Voice (Min)	TRX Voice (time)	Dou Voice (Day)	Payload Data (Gb)		
1	Low-early Internet adopter users	60.0%	9,017	11	6	4	2.0	Smartphone	A_1: Low Voice Usage Target
2	High-traditional users	4.8%	69,342	84	39	16	0.1	Non Smartphone	B: High Voice Usage Target
3	Internet user	9.2%	9,371	10	7	4	27.1	Smartphone	A_1: Low Voice Usage Target
4	Mid-traditional user	14.4%	16,648	22	10	6	0.2	Non Smartphone	A_2: Mid Voice Usage Target
5	Mid-early Internet adopter user	11.7%	59,892	73	33	14	4.0	Smartphone	B: High Voice Usage Target

Personalized Marketing

Mapping was conducted on the packages that were available or not available at Telkomsel in accordance with the segmentation that had been formed and the characteristics of behavioral targeting and customer positioning. According to Ozcelik & Kaan (2018), behavioral targeting performs really well only when consumers narrowly construe their preferences. In addition, Boerman, Kruikemeier & Borgeisus (2017) affirm that advertisers are increasingly monitoring people's online behavior and using the information collected to show people individually targeted advertisements. The following are the packages that could be offered to customers while still considering the concept of stimulating the usage of related customers.

Table 10 Product offer

Product						Campaign
A	Voice Personalize Offer • Based on ARPU segment (micro segment) • Offering about IDR 100-250 upper from ARPU • Voice quota adjusted to price and needs					
		Daily Voice ARPU	Example Price	Quota	Validity	
	A_1	<250	499	50 min	Daily	• SMS • Flash Message • Digital
		250-500	749	50 min	Daily	
		500-750	999	50 min	Daily	
		Etc.				
	A_2	1,000-2,000	2,499	100 min/UL	Daily	• SMS • Flash Message
		2,000-3,000	3,499	100 min/UL	Daily	
		Etc.				

Product					Campaign
B	BIG Quota Offer				
	• Monthly offering with Big Quota				
		Example			
	Package	Price	Quota	Validity	• SMS
	BIG Quota	70-100k	>1,000 min	Monthly	• Flash Message
	Unlimited	100-150k	UL min	Monthly	• Digital

Therefore, based on the obtained data above, the logistic regression results are obtained by selecting the most optimal production score to see the highest accuracy and sensitivity model, which is Iteration 2. These developmental models were then operationalized to predict the amount of highly prospect of voice package of total voice Payu users, which was 349,0006. The following formula of developmental model was used.

$$\pi = \frac{2.71828^{0.16-0.128Age-0.290Gender+\cdots\ldots\ldots+0Total\ Recharge}}{1+2.71828^{0.16-0.128Age+0.290Gender+\cdots\ldots\ldots+0Totalrecharge}}$$

Note:

π	:	Opportunity
α	:	Constant
$\beta 1....\beta n$	:	Regression coefficient for X1 ……. Xn variables
$X1....Xn$	:	Input variable X1 …….. Xn
e	:	Natural logarithm (2.71828)

Hence, it was revealed that the total number of highly prospective Telkomsel customers would buy a voice package of 144,342 from a total population of 349,006. Then the priority scale of 144,342 highly prospective customers who will buy a voice package was made by dividing it into five groups based on the purchasing score below:

Table X Prioritizing highly prospect customers

Group	Purchasing Score	Number of Customers	Numbers of Customers (%)
1	0.500 - 0.599	29,923	21%
2	0.600 - 0.699	18,834	13%
3	0.700 - 0.799	14,857	10%
4	0.800 - 0.899	14,513	10%
5	0.900 - 1.000	66,215	46%
Total		**144,342**	**100%**

It can be seen from the table above that there are 66,215 which are the top priority for highly prospects who will buy a voice package.

Discussion

Based on data processing results using the logistic regression algorithm model, a very good level of prediction accuracy was obtained. This means that if this prediction

model were applied to prepaid Telkomsel subscribers who had become voice package subscribers, which Voice PAYU subscribers could be predicted to become voice package subscribers. The evaluation was carried out with the measurement results for 97.2% accuracy and 98.2% sensitivity from the model formed. This is in line with the research conducted by Fudin (2018) which produced 79.7% accuracy in the purchasing prediction model using data mining to predict Telkomsel's prepaid customers based on five significant variables that influence customers to purchase data packages on My Telkomsel. Meanwhile, Noor (2018) found accuracy of 54.09%. It can be inferred that buying it or not, the calculation generated a higher value compared to previous research and the accuracy in this research shows the precision in predicting customers who would purchase voice packages. This is contrary to research by Shen (2014) that argues that preferences are constructed when decisions are being made, which means that customers often do not have stable preferences to be retrieved and applied to decisions. For instance, Ma et al. (2017) believed that when preferences often ill-defined, consumers will evaluate personalized recommendations based on how easily they can identify their stated preferences

In this research, the logistic regression model found seven significant variables affecting customer purchases from a total of 15 variables used. These variables can also be used to determine the strategy for making services according to customer characteristics, namely voice revenue, minutes of voice, voice transaction, payload data, SMS revenue, day of voice usage, and handset type. According to Kotler and Keller (2016), the market segment consists of a group of customers who have the same needs and the main task of a marketer is to identify which segment will be the target. The main factors or variables to create a market segment are geographic, demographic, psychographic, & behavioral factors (Kotler & Keller, 2016)

After the prediction model and clustering were formed, the next step was making a profile mapping according to the formed clusters associated with existing packages in Telkomsel or make new ones as needed. Five segments were formed and the clusters were named with reference to each cluster's characteristics and were categorized based on the order of each input variable used in the formation of the clustering model. The mapping was grouped into two main customer targets, namely low-mid voice usage segments and high-voice usage segments.

Seen from personalized marketing, the personalized marketing that can be offered to these two segments are personalized voice offers and big quota offer. These package offers are expected to increase the numbers of customer of voice package since they can get their desired products. The principal is similar to study by Lee & Chang (2011) that attitudes toward mass customization were generally positive, nearly half reported buying a personalized product, these buyers were very satisfied with their purchase.

There are some limitations of this research, in which it has not analyzed (1) prediction model like neural network and decision tree for better accuracy level, (2) segmentation method to know the cluster formation and approach of product offering to non voice package customers, and (3) the predictive analytics for customers who will churn to decide the personalized marketing. Therefore, for other researchers interested in conducted similar research, other predictive models such as neural networks and decision trees are recommended to be used to get a better level of accuracy.

Conclusion and Recommendation

Conclusion

Telkomsel responded to the declining trend of revenue starting 2014 by shifting their core business from Telephone and SMS services to Data and Digital services in 2018. Telephone service still contributed 28.4% to the revenue and was the second largest, while SMS gave 4.1%, hence, a prediction of voice package buyers is important to be identified to know the customer profiles and significant variables to form appropriate target customer segmentation

This research found a model and the prediction of voice PAYU customers who would buy voice packages using logistic regression algorithms which has an accuracy level of 97.2%. In addition, it is known that the total number of highly prospective Telkomsel customers who will purchase a voice package is 144,342 from a total population of 349,006. A priority scale was then made, in which 144,342 highly prospective customers will buy a voice package by dividing it into 5 groups based on the purchasing score. The clustering model result with a database set of 144,342 Telkomsel's highly prospective customers who would buy a voice package showed that 60% of the subscribers come from Low-Early Internet Adapter User segment with an average telephone usage in one month of 4 days with telephone usage, which reaches 11 minutes per month with voice ARPU per month of Rp. 9,017. In meaning that, obtaining good strategies is necessary to do build appropriate targeting and positioning. This can be done by mapping five segmentations divided into two main target customers with similar behavior, namely low-mid voice usage segment and high segments. Customers who are highly prospective to purchase voice packages are given two main types of offers based on customer profiles formed from five clustering results.

To obtain maximum expected results, the most important variables to make predictions to be considered are: the total average voice revenue in 1 month, average number of phone calls in 1 month (in minutes), number of telephone transactions in 1 month, total internet usage in 1 month (in Gb units), total SMS revenue in 1 month, number of days telephone usage within 1 month, the type of handset used by the customer. Hence, the performance of these variables are significant to be enhanced.

Recommendation

Based on the results of data analysis related to the predictive model & clustering model following the segmentation, targeting, and positioning, several suggestions are given for Telkomsel. First, the prediction model with the logistic regression method can be used as a tool for to predict customers who will purchase voice package services so that campaign activities that involve offering products to the customers can be more effective & efficient. Second, to find out the variables that significantly affect Telkomsel customers buying voice package services. The clustering model with K-Means can be used as a method of segmenting customers for Telkomsel to find out the similarity of the profiles of customers who will buy voice packages in one cluster and significantly differentiate between others clusters. Third, Telkomsel can also carry out behavioral targeting and product positioning based on the results of the prediction model & clustering model to be able to provide more targeted and right-on-target products.

Other segmentation methods to determine cluster formation and product offering approaches to non-voice package customers can also be considered other variables. To refine the analysis results in determining personalized marketing, predictive analytics for

customers who will churn is recommended to be analyzed. This research proposes recommendation for Telkomsel, as follows:

1. To use the prediction model using the logistic regression method to predict customers who want to buy voice package services, the campaign marketing activity can be more effective. The most significant variable influencing Telkomsel's customer preferences can be analyzed.
2. To use a clustering model with K-Means to segment the customer to know the similarity of customer profile who want to use voice package in a cluster, it can significantly differentiate other clusters.
3. To analyze behavioral targeting and product positioning based on the prediction model results and clustering models to provide more targeted products.

References

Boerman, S., Kruikemeier, S., & Borgesius, F. (2017). Online Behavioral Advertising: A Literature Review and Research Agenda. *Journal of Advertising*, 46, 363-376. 10.1080/00913367.2017.1339368.

Chaffey. D (2013), "Definitions of E-marketing vs Internet vs Digital marketing", Smart Insight Blog, February 16.

Dangi, H., & Malik, A. (2017). Personalisation in marketing: An exploratory study. *International Journal of Internet Marketing and Advertising*, 11., 124. Doi: 10.1504/IJIMA.2017.084079.

de Bellis, E., Hildebrand, C., Ito, K., Herrmann, A., & Schmitt, B. (2019). Personalizing the Customization Experience: A Matching Theory of Mass Customization Interfaces and Cultural Information Processing. *Journal of Marketing Research*, *56*(6), 1050-1065. Doi: https://doi.org/10.1177/0022243719867698

Ebrahim, R., Ghoneim, A., Irani, Z., & Fan, Y. (2016). A brand preference and repurchase intention model: the role of consumer experience. *Journal of Marketing Management*, *32*(13-14), 1230-1259. Doi: /https://doi.org/10.1080/0267257X.2016.1150322

Franedya, R. (2019, May 01). Industri Telco Tumbuh Negatif, Ini Strategi Bisnis Telkomsel [Telco Industry Grows Negatively, This Is Telkomsel's Business Strategy]. Retrieved from: https://www.cnbcindonesia.com/tech/20190501182321-37-69945/industri-telco-tumbuh-negatif-ini-strategi-bisnis-telkomsel

Fudin, U. (2018). Meningkatkan Jumlah Pembeli Paket data Di MyTelkomsel Apps Dengan Predictive Analytics [Increasing the NUMBER of Data Package Buyers on MyTelkomsel Apps with Predictive Analytics]. Unpublished Thesis: Telkom University. Retrieved from: https://openlibrary.telkomuniversity.ac.id/home/catalog/id/147181/slug/meningkatkan-jumlah-pembeli-paket-data-di-mytelkomsel-apps-dengan-predictive-analytics.html

Goldsmith, R. E. (1999). The personalised marketplace: beyond the 4Ps. *Marketing Intelligence & Planning*.

Hjort, K., Lantz, B., Ericsson, D., & Gattorna, J. (2016). Customer Segmentation Based on Buying and Returning Behaviour: Supporting Differentiated Service Delivery

in Fashion E-Commerce. In *Developments in Logistics and Supply Chain Management* (pp. 153-169). Palgrave Macmillan, London.

Hosmer Jr, D. W., Lemeshow, S., & Sturdivant, R. X. (2013). *Applied logistic regression* (Vol. 398). John Wiley & Sons.

Kotu, V. & Deshpande, B. (2015) *Predictive Analytics and Data Mining, Business Intelligence.* Waltham: Morgan Kaufmann Publishers.

Kotler, P., & Armstrong, G. (2014). *Principles of Marketing*. New Jersey: Prentice Hall.

Kotler, P., & Lane, K (2016). *Marketing Management (Global Edition ed.).* USA: Pearson Education Limited. ISBN 978-0-13-385646-0

Kurniawan, S. (2019). Increasing Number of Active Data User in Telkomsel Using Predictive Analysis. Unpublished Thesis: Telkom University.

Lee, H. H., & Chang, E. (2011). Consumer Attitudes Toward Online Mass Customization: An Application of Extended Technology Acceptance Model. *Journal of Computer-Mediated Communication*, *16*(2), 171-200. Doi: https://doi.org/10.1111/j.1083-6101.2010.01530.x

Lin, W., Ke, S., & Tsai, C. (2017). Top 10 data mining techniques in business applications: a brief survey. *Kybernetes, 46*, 1158-1170. Doi : https://doi.org/10.1108/K-10-2016-0302

Ma, Y., Chen, G., & Wei, Q. (2017). Finding users preferences from large-scale online reviews for personalized recommendation. *Electronic Commerce Research*, *17*(1), 3-29. https://doi.org/10.1007/s10660-016-9240-9

Nisbet, R., Elder, J., & Miner, G. (2009). *Handbook of statistical analysis and data mining applications*. Academic Press.

Noor, I. G. W. A., Ariyanti, M., & Alamsyah, A. (2019, May). Telecom Customer's Segmentation Using Decision Tree to Increase Active Electronic Money Subscribers. In *1st International Conference on Economics, Business, Entrepreneurship, and Finance (ICEBEF 2018)* (pp. 628-632). Atlantis Press.

Ozcelik, A & Varnali, K. (2018). Effectiveness of online behavioral targeting: a psychological perspective. *Electronic Commerce Research and Applications*, *33*, 100819. Doi: 10.1016/j.elerap.2018.11.006.

Paço, A & Alves, H.,Shiel, C & Filho, W. (2014). An analysis of the measurement of the construct "buying behaviour" in green marketing. *Journal of Integrative Environmental Sciences*, *11*(1), 55-69. Doi: 11. 10.1080/1943815X.2014.894082.

Pal, J., & Bhattacherjee, V. (2015). Fuzzy Rule Based Classifier for Software Quality Data. *Journal of Theoretical and Applied Information Technology*, *72*(1), 58-66. ISSN: 1992-8645.

Shen, A. (2014). Recommendations as personalized marketing: insights from customer experiences. *Journal of Services Marketing*, *28*(5), 414-427. Doi: 10.1108/JSM-04-2013-0083.

Swastha, B., & Irawan. (2005). *Asas-asas Marketing*. Liberty: Yogyakarta.

Telkomsel Annual Report. (2019). Retrieved from https://telkomsel.com/about-us/investor-relations

Tuckwell, K. J. (2016). *Think Marketing.* Canada: Pearson Canada Inc.

Vesanen, J. (2005). What is personalization?: a literature review and framework.

Wendler, T., & Gröttrup, S. (2016). *Data mining with SPSS modeler: theory, exercises and solutions*. Springer.

Widiyaningtyas, T., Prabowo, M. I. W., & Pratama, M. A. M. (2017, September). Implementation of K-means clustering method to distribution of high school

teachers. In *2017 4th International Conference on Electrical Engineering, Computer Science and Informatics (EECSI)* (pp. 1-6). IEEE. Doi: https://doi.org/10.1109/EECSI.2017.8239083