Mobile Customer Behavior Predictive Analysis for Targeting Netflix Potential Customer

Suryadi Tanuwijaya[1], Andry Alamsyah[2], Maya Ariyanti[3]

[1, 2, 3] School of Economic and Business, Telkom University
stanuwijaya@student.telkomuniversity.ac.id[1], andrya@telkomuniversity.ac.id[2], ariyanti@telkomuniversity.ac.id[3]

Abstract

The development of Indonesia ICT environment has made mobile video-on-demand (VOD) platform as one of emerging lifestyle. With advance smartphone technology, mobile phone subscribers able to enjoy high resolution mobile VOD service with greater user experience. The purpose of this study is to profile and predict potential customer of one of VOD platform, Netflix, for personalize marketing target. Using machine learning predictive analytic methodology, customer profile and behavior data is divided into 3 clusters using K-Means model before tested with several supervised model for getting best model for each cluster. Feature importance analysis will give marketing insight for product offering follows up to each targeted potential customer. Significant variables affecting Netflix buyer and non-buyer within 3 different clusters are defined clearly with number of potential customer that can be targeted as future subscriber of Netflix. Based on the research results, this method can be used by mobile operator to target potential customer with effective promotional or product offering by personalized marketing approach based on behavioral pattern and customer needs. It is expected by implementing this methodology, effectivity and accuracy of marketing will be increased compared to conventional method.

Keywords: Predictive Analytic, behavior, Personalized Marketing.

1. Introduction

Emerging internet services in Indonesia lately has created huge digital service market opportunity. With fourth largest population in world, Indonesia online video and music streaming services is predicted to reach USD 10 billion revenue with 18% CAGR by 2025 [1]. Video service is currently dominating internet traffic portion, it's predicted that carried mobile video traffic will become 74% of total cellular traffic by 2024 [2]. While mobile operator this far only provides connectivity service, and its broadband revenue unable to catch up business growth target, most of cellular operators are now transforming into digital company instead of network provider. Collaboration with service provider or OTT apps will become key in able to compete in digital era. While operator can maintain quality of service to customer for each service from OTT platform, partnership between operator and OTT, including VOD provider which becomes inevitability within this digital era [3]. Netflix, one of most popular VOD platform in global and Indonesia, is chosen as object for this research.

This research is aimed to develop predictive analytic methodology for detecting potential subscriber of Netflix from mobile customer profile and behavior data on historical mobile app usage transaction. Besides getting potential subscriber number, it is intended to output profile and customer who has subscribed to Netflix. Prediction is being done by using combination of K-Means clustering and 7 classification models which being tested to each cluster. In the end, to formulate personalization marketing plan, profile of potential subscriber and actual Netflix subscriber are to be compared and analyzed to see which apps can be used as targeting channel to potential subscriber. Dataset that used in this research are collected during September – December 2019, before COVID-19 pandemic.

2. Related Work

There are prior work on studying predictive analysis using big large dataset to study customer behavior for marketing purpose. While studying consumer behavior from customer interaction in online platform is getting more prospectus in digital era [4] some relevant research have been conducted by using similar methodology. Conceptual approach on using predictive analytics with behavior informatics and analytics approach is examined in [5], where historical behavioral data from customer transaction can be used to get deeper understanding on consumer behavior in order to make better marketing campaign. Chen et al. [6] conducted customer behavior analysis based on mobile app usage to foresee high value customer for further app development. While mobile customer data has been used for predicting potential churn [7] and prospective customer of electronic money application [8] both using classification supervised model.

Segmentation modeling based on customer usage on voice and data usage along with customer spending has studied in [9], [10]. While all mentioned research were conducted from network provider point of view, Rahman et al. [11] examined video consumer behavior from OTT provider perspective. Based on personalization marketing framework in [12], this research will covering scope from customer data and external data as input until customization process.

. 3. Research Methodology

The research is done following research stage as shown in Figure 1. Mobile data subscriber from 6 big cities in Indonesia chosen as research object and 4 months data which is during September – December 2019, before COVID-19 took place at early of 2020. Research stage and data mining process adopted CRISP-DM model [13].

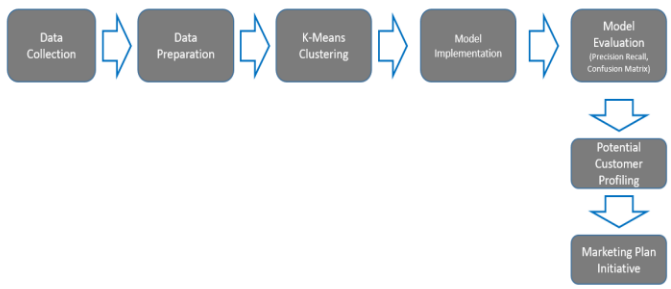

Figure 1. Research Stage

A. Datasets

Collected data is based on customer level as unit with monthly basis. It is represent with unique ID for each customer. Datasets is combined from total 139 variables which consists of Customer ID with encrypted MSISDN, age, gender and Smartphone brand.

Behavior variables, represents with transaction on 45 popular mobile apps which represents with traffic, duration and session features. Mobile apps are group into several category based on their service. Session is number of accessed conducted to certain mobile apps, while duration is how long time consumed in every access to mobile apps, while traffic is volume of traffic carried in every access to mobile apps.

Table 1 Mobile Apps Category

Category	***Number of apps***
Video Streaming	6
Music Streaming	5
Games	5
E-Wallet	5
File Transfer	5
E-Banking	4
E-Commerce	5
Instant Messaging	5
VOIP	5
TOTAL APPS	45

B. Preparation, Downsampling and Clustering

Data preparation process takes place right after all necessary data collected from sources. Data preparation usually takes longer time compared to other stages. Basically there are few important things that being done in data preparation stages.

- Data cleaning, substitution and reparation.
- New variables based on group or classification
- Netflix subscriber and non-subscriber flag based on traffic, session and duration variables.

Since it is found out that Netflix user only acquired 2.6% from total population, which will create imbalance dataset, under sampling method is preferred compare to oversampling method [14] before processed further into clustering stage.

Table 2 Downsampling Dataset

Status Buy	Description	Subs	% Subs
1	Netflix Subscriber	1.056.582	47%
0	Netflix Non-Subscriber	1.172.824	53%

In order to divide dataset into smaller groups with similar characteristic, K-Means model chosen as clustering methodology. By using elbow method, it is determined that dataset is split into 3 cluster which divided dataset group characteristic into Low, Medium and Heavy user based on traffic, session and duration of mobile app historical transaction usage as describer in Table III. Before entering modeling process, dataset divided into training and testing dataset with holdout validation method by 75% and 25% respectively in each cluster.

Table 3 Clustering Result

Feature	Cluster 1	Cluster 2	Cluster 3
Duration (seconds)	57.27	1332.83	434.36
Session	5.85	49.35	26.54
Traffic (mb)	24.74	598.02	223.5

C. Modeling and Evaluation

With 3 different cluster which represent different level of customer consumption on data transaction based on mobile app usage history, several classification models are implemented and tested for getting best model that fit in every cluster. Performance evaluation conducted by determining model accuracy and sensitivity from confusion matrix value. There are 7 supervised machine learning and ensemble model are tested:

1) Naïve Bayes
2) Decision Tree
3) Random Forest
4) Logistics Regression
5) LGBM
6) XGBoost
7) Catboost

	Metode	Accuracy_test	Precision_test	Recall_test	F1_test	AUC_test
6	Catboost	0.836787	0.748064	0.871174	0.871174	0.843146
4	XGBoost	0.820320	0.718347	0.859442	0.859442	0.828081
5	LGBM	0.811260	0.707947	0.847682	0.847682	0.818499
2	Random Forest	0.805683	0.728202	0.819997	0.819997	0.808073
0	Decision Tree	0.728173	0.699821	0.697401	0.697401	0.725464
3	Logistic Regression	0.685794	0.733845	0.629554	0.629554	0.688723
1	Naive Bayes	0.589132	0.139125	0.728634	0.728634	0.652324

Figure 2 Model Evaluation Cluster 1

	Metode	Accuracy_test	Precision_test	Recall_test	F1_test	AUC_test
6	Catboost	0.843950	0.889823	0.879468	0.879468	0.823941
5	LGBM	0.842631	0.887300	0.879655	0.879655	0.822146
4	XGBoost	0.838673	0.884497	0.876632	0.876632	0.817637
2	Random Forest	0.796645	0.876367	0.830499	0.830499	0.772254
0	Decision Tree	0.724651	0.791982	0.797122	0.797122	0.688005
3	Logistic Regression	0.692612	0.950378	0.699835	0.699835	0.658359
1	Naive Bayes	0.398982	0.135688	0.820339	0.820339	0.583304

Figure 3 Model Evaluation Cluster 2

	Metode	Accuracy_test	Precision_test	Recall_test	F1_test	AUC_test
6	Catboost	0.866384	0.890304	0.905470	0.905470	0.849898
4	XGBoost	0.856806	0.886058	0.895657	0.895657	0.839788
5	LGBM	0.847696	0.878947	0.888832	0.888832	0.829722
2	Random Forest	0.814208	0.868668	0.852381	0.852381	0.794196
0	Decision Tree	0.741723	0.801604	0.805704	0.805704	0.712704
3	Logistic Regression	0.699169	0.939204	0.703709	0.703709	0.684787
1	Naive Bayes	0.405360	0.121929	0.836919	0.836919	0.598208

Figure 4 Model Evaluation Cluster 3

Based on evaluation on 3 clusters, it shows that Catboost is the best model on every clusters with highest accuracy and sensitivity or recall among all tested models.

4. Results

After model evaluation done and Catboost model is selected as best performed model on all clusters, feature importance can be output to determine significant variables on both Netflix subscriber and non-subscriber. Predictive analysis on potential subscriber who might want to subscribe to Netflix while in actual not yet, also can be output within this stage. Based on features importance analysis from Netflix subscriber, top significant apps that related with customer who subscribed to Netflix are Vidio (Video Streaming), Spotify (Music Streaming) and Garena (Games). Itune file access apps describes that Iphone users are also significant factor, followed by 6 E-commerce apps which shows customer who is actively using E-commerce has big chance on subscribing to Netflix.

Table 4 Feature Importance on Netflix Subscriber

Category	Apps	Importance
Video	Vidio	5.7
Music	Spotify	4.44
Games	Garena	6.56
File Access	Itune	3.18
E-Wallet	Ovo	4.88
E-Commerce	Amazon	6.6
E-Commerce	Lazada	4.62
E-Commerce	Tokopedia	4.21
E-Commerce	Shopee	4.93
E-Commerce	Bukalapak	3.52
E-Commerce	Jd.Id	3.45

Based on prediction result using Catboost model, number of potential customer who is predicted to do subscription but in actual are still not yet subscribed represented with False Positive (FP) number on confusion matrix. 3 different group of potential customer taken from cluster 1, 2 and 3 with different consumption characteristic which might impact on targeting priority later on during acquisition process. Total potential customer is 22,892 subscriber. As described in Table V, highest number 19,699 in cluster 1 with low consumption user which will be last priority to target.

Table 5 Number of Potential Subscriber

Cluster	FP (Potential Subscriber)	% (FP/ Non-subscriber)
1 (Low Consumption)	19,699	9.06%
2 (High Consumption)	435	25.01%
3 (Medium Consumption)	2,758	17.99%
Total	22,892	9.76%

With information on feature importance profile and customer can be classified as potential subscriber of Netflix, next thing before personalized marketing plan formed is comparing profile on actual subscriber with potential subscriber. Subscriber with False Positive classification is our potential customer to be targeted, then to be compared with actual subscriber of Netflix, based on significant variables on each cluster. Identical significant app is taken as variable comparison between potential and actual subscriber. Tabel 6 shows result of app comparison along with gap analysis.

Table 6 Actual and Potential Customer Profile Comparison

Cluster	Category	Feature	App name	Actual Importance	Potential Importance
1	E-Commerce	duration	shopee	528.22	85.08
1	File Access	trafficmb	itune	46.17	21.69
1	E-Commerce	session	tokopedia	43.17	22.39
2	Games	session	garena	1314.7	194.37
2	E-Wallet	session	ovo	122.68	24.6
2	E-Commerce	session	jd.id	166.3	75.6
2	Video	trafficmb	vidio	58.11	30.07
3	E-Wallet	session	ovo	1015.9	208.2
3	E-Commerce	duration	shopee	135.2	36.63
3	File Access	session	itune	85.36	47.23

Based on analysis provided, every cluster can be predicted and targeted with different product offering based on described profiler on each cluster.

1) Cluster 1 or low consumption Cluster, doesn't subscribe on VOD platform instead only using free VOD platform like Youtube with low access frequency and duration. Active in using E-Commerce platform like Shopee and Tokopedia. iPhone users within this cluster shows high similarity with Netflix subscriber.

2) Cluster 2 or high consumption cluster, subscribed to VOD platform which is Vidio. High frequent in using E-Commerce platform like Jd.id and OVO as E-Wallet payment platform. Detected as heavy gamer with Garena apps.

3) Cluster 3 or medium consumption cluster, has similarity in profile with Cluster 2 in high frequent of E-Commerce Shopee platform with OVO as payment platform. iPhone user within this cluster can be prioritized for initial targeting since shows high similarity with Netflix subscriber.

Following predictive analysis result which has divided subscriber based on behavior in consumption into 3 clusters, determine which customer are potential to be acquired further with marketing targeting, and profile similarity between potential subscriber and actual Netflix subscriber, marketing plan can be formed with different product offering and different priority in acquisition. As per shown in Figure 5, product divided into 3 different cluster based on its each profile. Cluster 1 as low consumption data user, can be offered with low volume and amount of data package with low resolution on Netflix for 1 month. Bundling voucher with frequent used E-Commerce platform can be included in personalized marketing program for cluster 1 subscribers. While Cluster 2 with higher spending in data consumption and behavior can be offered with premium package with high quality on video resolution and combined game package. Cluster 2 should be prioritized in upfront considering monetary factor that can beneficial to mobile operator. Cluster 3 as medium cluster, need to be uplift so it can be similar like Cluster 2. Medium volume data package and 720p resolution can be offered to cluster 2 subscriber as trigger for future uplift in data usage or Netflix package subscription.

5. Conclusion

Predictive analysis using mobile customer behavioral data can be implemented for use in supporting mobile operator daily business in supporting marketing and sales targeting to potential customer. By doing this method it is expected campaign cost can be reduced and deliver the right product for the right customer. Same methodology can be used for other VOD platform aside Netflix.

References

[1] Google, Temasek and Bain & Company. (2019, October). e-Conomy SEA 2019 report. https://www.blog.google/documents/47/SEA_Internet_Economy_Report_2019.pdf

[2] Ericsson. (2020, November). Ericsson Mobility Report. Fredrik Jejdling. https://www.ericsson.com/4adc87/assets/local/mobility-report/documents/2020/november-2020-ericsson-mobility-report.pdf

[3] Minzheong, S. (2020). A Case Study on Partnership Types between Network Operators & Netflix: Based on Corporate Investment Model. International Journal of Internet, Broadcasting and Communication, 12(1), 14–26. https://doi.org/10.7236/IJIBC.2020.12.1.14

[4] Hofacker, C.F., Malthouse, E.C. and Sultan, F. (2016), "Big Data and consumer behavior: imminent opportunities", Journal of Consumer Marketing, Vol. 33 No. 2, pp. 89-97. https://doi.org/10.1108/JCM-04-2015-1399

[5] Asniar and K. Surendro, "Predictive Analytics for Predicting Customer Behavior," 2019 International Conference of Artificial Intelligence and Information Technology (ICAIIT), Yogyakarta, Indonesia, 2019, pp. 230-233, doi: 10.1109/ICAIIT.2019.8834571.

[6] Chen, Q., Zhang, M. and Zhao, X. (2017), "Analysing customer behaviour in mobile app usage", Industrial Management & Data Systems, Vol. 117 No. 2, pp. 425-438. https://doi.org/10.1108/IMDS-04-2016-0141

[7] Lee, E.-B., Kim, J. and Lee, S.-G. (2017), "Predicting customer churn in mobile industry using data mining technology", Industrial Management & Data Systems, Vol. 117 No. 1, pp. 90-109. https://doi.org/10.1108/IMDS-12-2015-0509

[8] Noor, I & Ariyanti, Maya & Alamsyah, Andry. (2019). Telecom Customer's Segmentation Using Decision Tree to Increase Active Electronic Money Subscribers. 10.2991/icebef-18.2019.134.

[9] S. Aheleroff, "Customer segmentation for a mobile telecommunications company based on service usage behavior," The 3rd International Conference on Data Mining and Intelligent Information Technology Applications, Macao, 2011, pp. 308-313.

[10] Widodo, A., & Ramantoko, G. (2019). Optimising Market Segmentation for The Telecommunications Industry: A Contextual Marketing Based Approach: Case Study at PT TELKOMSEL. Asian Journal of Management Sciences & Education, 8(1), 57–65.

[11] S. Rahman, H. Mun, H. Lee, Y. Lee, M. Tornatore and B. Mukherjee, "Insights from Analysis of Video Streaming Data to Improve Resource Management," 2018 IEEE 7th International Conference on Cloud Networking (CloudNet), Tokyo, 2018, pp. 1-3, doi: 10.1109/CloudNet.2018.8549180.

[12] Jari Vesanen, (2007),"What is personalization? A conceptual framework", European Journal of Marketing, Vol. 41 Iss 5/6 pp. 409 – 418.

[13] Azevedo, Ana & Santos, Manuel. (2008). KDD, semma and CRISP-DM: A parallel overview. 182-185.

[14] Drummond, C. (2003). C 4 . 5 , Class Imbalance , and Cost Sensitivity : Why Under-Sampling beats OverSampling.