



## RESEARCH ARTICLE

# Customer churn prediction for telecommunication industry: A Malaysian Case Study [version 1; peer review: 2 approved]

Nurulhuda Mustafa<sup>1</sup>, Lew Sook Ling <sup>2</sup>, Siti Fatimah Abdul Razak <sup>2</sup>

<sup>1</sup>Telekom Malaysia, Faculty of Business, Ayer Keroh, Melaka, 75450, Malaysia

<sup>2</sup>Faculty of Information Science and Technology, Multimedia University, Ayer Keroh, Melaka, 75450, Malaysia

**V1** First published: 13 Dec 2021, 10:1274  
<https://doi.org/10.12688/f1000research.73597.1>  
 Latest published: 13 Dec 2021, 10:1274  
<https://doi.org/10.12688/f1000research.73597.1>

## Abstract

**Background:** Customer churn is a term that refers to the rate at which customers leave the business. Churn could be due to various factors, including switching to a competitor, cancelling their subscription because of poor customer service, or discontinuing all contact with a brand due to insufficient touchpoints. Long-term relationships with customers are more effective than trying to attract new customers. A rise of 5% in customer satisfaction is followed by a 95% increase in sales. By analysing past behaviour, companies can anticipate future revenue. This article will look at which variables in the Net Promoter Score (NPS) dataset influence customer churn in Malaysia's telecommunications industry.

The aim of This study was to identify the factors behind customer churn and propose a churn prediction framework currently lacking in the telecommunications industry.

**Methods:** This study applied data mining techniques to the NPS dataset from a Malaysian telecommunications company in September 2019 and September 2020, analysing 7776 records with 30 fields to determine which variables were significant for the churn prediction model. We developed a propensity for customer churn using the Logistic Regression, Linear Discriminant Analysis, K-Nearest Neighbours Classifier, Classification and Regression Trees (CART), Gaussian Naïve Bayes, and Support Vector Machine using 33 variables.

**Results:** Customer churn is elevated for customers with a low NPS. However, an immediate helpdesk can act as a neutral party to ensure that the customer needs are met and to determine an employee's ability to obtain customer satisfaction.

**Conclusions:** It can be concluded that CART has the most accurate churn prediction (98%). However, the research is prohibited from accessing personal customer information under Malaysia's data protection policy. Results are expected for other businesses to measure potential customer churn using NPS scores to gather customer feedback.

## Open Peer Review

Approval Status

|                  | 1                    | 2                    |
|------------------|----------------------|----------------------|
| <b>version 1</b> |                      |                      |
| 13 Dec 2021      | <a href="#">view</a> | <a href="#">view</a> |

1. **Catur Supriyanto** , Universitas Dian Nuswantoro, Semarang, Indonesia
2. **D. I. George Amalarethinam** , Jamal Mohamed College, Tiruchirappalli, India

Any reports and responses or comments on the article can be found at the end of the article.

**Keywords**

Customer Churn, Net Promoter Score (NPS), Data Mining Techniques, Classification and Regression Trees (CART)



This article is included in the **Research Synergy** Foundation gateway.

**Corresponding author:** Lew Sook Ling ([sllew@mmu.edu.my](mailto:sllew@mmu.edu.my))

**Author roles:** **Mustafa N:** Data Curation, Investigation, Methodology, Software, Validation, Visualization, Writing – Original Draft Preparation; **Sook Ling L:** Conceptualization, Methodology, Project Administration, Supervision, Validation, Writing – Review & Editing; **Abdul Razak SF:** Conceptualization, Methodology, Software, Supervision, Validation, Writing – Review & Editing

**Competing interests:** No competing interests were disclosed.

**Grant information:** The author(s) declared that no grants were involved in supporting this work.

**Copyright:** © 2021 Mustafa N *et al.* This is an open access article distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

**How to cite this article:** Mustafa N, Sook Ling L and Abdul Razak SF. **Customer churn prediction for telecommunication industry: A Malaysian Case Study [version 1; peer review: 2 approved]** F1000Research 2021, 10:1274 <https://doi.org/10.12688/f1000research.73597.1>

**First published:** 13 Dec 2021, 10:1274 <https://doi.org/10.12688/f1000research.73597.1>

## Introduction

Customer retention and customer satisfaction are essential for a business to succeed.<sup>1</sup> Customer satisfaction is improved by repeating businesses, brand loyalty, and positive word of mouth.<sup>2</sup> Consumers prefer to stay with their current providers due to quality and price. Therefore, new anti-churn strategies must be constantly developed.<sup>3</sup> Data processing automates analytical model building. Machine learning algorithms improve the dataset iteratively to find hidden patterns.<sup>4</sup> Several studies show that machine learning can predict churn and severe problems in competitive service sectors. Predicting churning customers early on can be a valuable revenue source.<sup>5</sup> The results of the mediating effects of a customer's partial defection on the relationship between churn determinants and total defection show that some churn determinants influence customer churn, either directly or indirectly through a customer's status change, or both; thus, a customer's status change explains the relationship between churn determinants and the probability of churn.<sup>6</sup> This study hypothesised that changes in Net Promoter Score (NPS) can indicate whether churn determinants directly or indirectly influence churn.

## Telecommunication industry and service providers in Malaysia

Figure 1 presents the revenue generated by telecommunications in Malaysia. Telekom Malaysia (TM) has made RM11.43 billion in 2019 and ranked top.<sup>7</sup>

Malaysia's mobile, fixed broadband, and household penetration are expected to grow further between 2019 and 2025, providing 30 to 500 Mbps fibre internet for cities, with gigabit connections for industries.<sup>8</sup> As a result, customers expect the same or better service from providers. This statistic is also used to compare a company's performance to competitors.<sup>9</sup>

## Churn rates in Malaysia

Total customer turnover is the number of customers leaving the provider.<sup>10</sup> In the telecommunications industry, market competitiveness is measured by churn rate. Telephone, internet, and mobile services are all part of telecommunications. For example, one of the Internet Service Provider (ISP) from 20 subscribers will cancel, reducing annual revenue by 5%.<sup>11</sup> Every day, the telecom industry loses 20%-40% of its customers.<sup>12</sup> Without pricing and subscription plans, 83% of Malaysians would switch telecom providers. On the other hand, 66% had no problem cancelling their existing service provider subscription.<sup>13</sup> Customer satisfaction metrics would help providers to sustain their customers.

## Net Promoter Score (NPS)

Net Promoter Score (NPS) measures customer satisfaction and loyalty using a 10-point Likert scale.<sup>14</sup> Consistency in purchases shows commitment regardless of performance and fulfilment.<sup>15</sup> Risk scores are used to predict customer churn. Churn is predicted using customer profiles and transaction patterns—predictive analytics use demographic, transactional data and NPS.<sup>16</sup> The NPS outperforms customer satisfaction.<sup>17</sup> A high NPS indicates that word of mouth can help businesses thrive. Customers are classified as promoters, detractors, or passives after receiving assistance from the helpdesk (Table 1).

## Customer satisfaction

Customer satisfaction is defined as customers being happy with a company's products, services, and capabilities.<sup>18</sup> Customers' satisfaction is influenced by buyer experience.<sup>19</sup> Satisfied customers lead to more sales and referrals. Proactive personal helpdesk and staff assistance require a company's ability to anticipate customer needs. Most businesses benefit from happy customers. In this study, the NPS measures customer satisfaction.

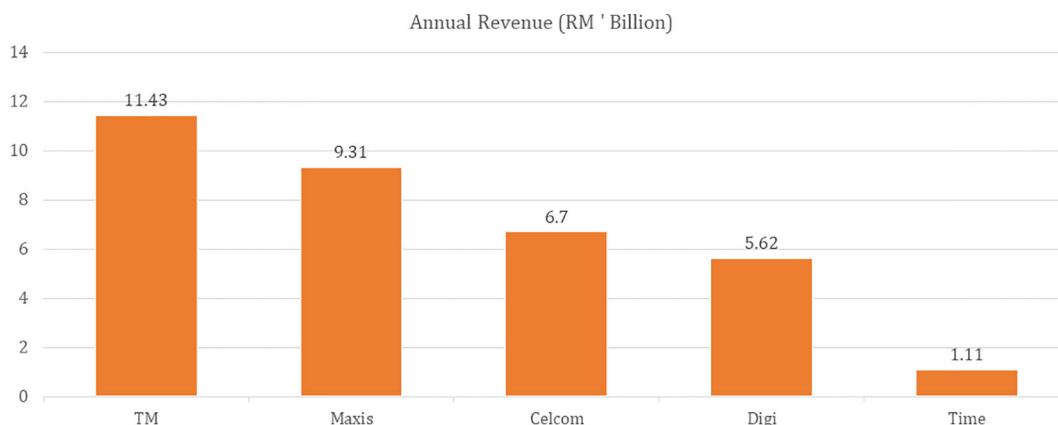


Figure 1. Telecommunications and Internet industry: Participants Annual Revenue as of 31<sup>st</sup> December 2019.

**Table 1. Net Promoter (NPS) scale.**

| Scale      | Score | Description   |
|------------|-------|---|
| Promoters  | 9-10  | Customers who are typically the brand's ambassadors, enhancing a brand's reputation and/publicity and referrals flow.                               |
| Passives   | 7-8   | Customers who have positive/constructive feelings towards the brand but are not expressing a need to change.  |
| Detractors | 0-6   | Customers who are unlikely to remain or encourage others to return—and even worse—may discourage others from trying to trust the business or brand. |

Source: 2021 Guide &amp; Definition (2021).

**Table 2. A summary of churn prediction studies.**

| Author (Year)              | Techniques and method  | The disadvantage of the prediction studies & proposed enhancement   |
|----------------------------|--|---|
| Ahmad et al. (2019)        | <b>Techniques:</b><br>Decision Tree, Random Forest, Gradient Boost Machine Tree, and XGBoost.<br><b>Method:</b><br>Churn predictive system using Hortonworks Data Platform (HDP) categorised under specific specialisation like Data Management, Data Access, Security, Operations and Governance Integration. <sup>22</sup> | <b>Disadvantage:</b><br>Only about 5% of the dataset entries represent customer's churn.<br><b>Proposed enhancement:</b><br>The dataset entries can be solved using under-sampling or trees algorithms.   |
| Höppner et al. (2020)      | <b>Techniques:</b><br>Decision Tree<br><b>Method:</b><br>ProfTree learns profit-driven decision trees using an evolutionary algorithm. ProfTree outperforms traditional accuracy-driven tree-based methods in a benchmark study using real-life data from various telecommunication service providers. <sup>23</sup>         | <b>Disadvantage:</b><br>The evolutionary algorithm (EA) training times are relatively slower than other classification algorithms.<br><b>Proposed enhancement:</b><br>Combining the ProfTree algorithm with random forests further optimises property by creating profit-driven trees and aggregating them.   |
| Yang (2019)                | <b>Techniques:</b><br>Random Forest<br><b>Method:</b><br>A new T+2 churn customer prediction model was proposed, in which churn customers are identified in two months, and a one-month window T+1 is set aside for implementing churn management strategies. <sup>24</sup>  | <b>Disadvantage:</b><br>In the T+2 churn prediction, a precision ratio of about 50% was achieved, with a recall ratio of about 50%.<br><b>Proposed enhancement:</b><br>Proposed to use more than one algorithm technique to compare the outcome result.   |
| Ahmed and Maheswari (2017) | <b>Techniques:</b><br>Firefly<br><b>Method:</b><br>Simulated Annealing modifies the actual firefly algorithm comparison to provide faster and more accurate churn predictions. <sup>25</sup>   | <b>Disadvantage:</b><br>The proposed algorithm's accuracy is comparable to that of the average firefly algorithm.<br><b>Proposed enhancement:</b><br>Incorporation of schemes or modifications to reduce False Positive rates.  |
| Eria and Marikannan (2018) | <b>Techniques:</b><br>Support Vector Machines, Naive Bayes, Decision Trees, and Neural Networks.<br><b>Method:</b><br>This study looks at 30 CCP studies from 2014-2017. Some data preparation and churn prediction issues arose. <sup>26</sup>  | <b>Disadvantage:</b><br>Telecom datasets should be handled cautiously due to their unbalanced nature, large volume, and complex structure.<br><b>Proposed enhancement:</b><br>It is necessary to investigate a gap in real-time churn prediction using big data technologies. More adaptable approaches that consider changing national or regional economic conditions are required. Customer churn could also be identified using word and voice recognition. |

**Table 3. A summary for churn prediction framework studies.**

| Author (Year)                 | Clemes et al. (2010)   | Geetha and Abitha Kumari (2012)   | Kim et al. (2017)   |
|-------------------------------|--|---|---|
| <b>Ahn et al. (2006)</b>      | Identifies and analyses factors influencing bank customers' switching behaviour in the Chinese retail banking industry. <sup>28</sup>  | Provides a brief overview of the trend of non-revenue earning customers (NRECs) that trigger revenue churn and are likely to churn soon. <sup>29</sup>  | Analyzes the factors that are affecting IPTV service providers' behaviour regarding switching barriers, VOCs, and content consumption. <sup>30</sup>  |
| <b>Propose</b>                | Describes a customer's status transition from active to non-user or suspended as a partial defection and from functional to absolute defective defection from active to churn. <sup>27</sup>   |   |   |
| <b>Dataset and techniques</b> | <b>Dataset:</b> 10,000 random samples from leading providers of mobile telecommunications services in South Korea<br><b>Technique:</b> Logistic regression   | <b>Dataset:</b> 37,388 datasets from a leading telecom service Provider<br><b>Technique:</b> Linear modelling   | <b>Dataset:</b> 5000 datasets from IPTV users in South Korea/<br><b>Technique:</b> Logistic regression  |
| <b>Variables</b>              | <b>Dependent variables</b><br>Customer churn<br><b>Independent variables</b><br>Customer Dissatisfaction<br>Call drop rate<br>Call failure rate<br>Number of complaints<br>Switching costs<br>Loyalty points<br>Membership card<br>Service usage<br>Billed amounts<br>Unpaid balances<br>Number of unpaid monthly bills<br>Customer status<br>Customer-related variables | <b>Dependent variables</b><br>Susceptibility to Churn<br><b>Independent variables</b><br>The extent of Local Calls and STD Calls to Other Networks<br>Multiple directory numbers<br>Rate plan<br>Tariff<br>Admin fee<br>Count of recharge<br>Sum of recharge<br>VAS Usage<br>Total minutes of usage VAS<br>Overall usage revenue per minute slab<br>Total usage revenue | <b>Dependent variables</b><br>Customer Behaviours<br>Customer defection<br><b>Independent variables</b><br>Switch Barrier<br>Service bundling<br>Remaining contract months<br>Membership points<br>VOC (Voice of Customer)<br>Total VOC<br>Membership Period<br>Membership Years<br>Degree of Content Usage<br>Channel views<br>VOD views<br>PPV<br>Monthly subscriptions |

**Table 3.** Continued

| Author (Year)  | Clemes et al. (2010)   | Geetha and Abitha Kumari (2012)   | Kim et al. (2017)  |
|--|--|---|--|
| <p><b>Ahn et al. (2006)</b></p> <p><b>Hypotheses</b></p> <p><b>H1</b> Customer Dissatisfaction à Customer churn<br/> <b>H2</b> Switching costs à Customer churn<br/> <b>H3</b> Service usage à Customer churn<br/> <b>H4</b> Customer status à Customer churn<br/> <b>Mediation effects</b><br/> <b>H1'</b> Customer Dissatisfaction à Customer Status<br/> <b>H2'</b> Switching costs à Customer Status<br/> <b>H3'</b> Service usage à Customer Status</p> | <p><b>H1</b> price à customers switching banks<br/> <b>H2</b> reputation à customers switching banks<br/> <b>H3</b> service quality à customers switching banks<br/> <b>H4</b> Effective advertising by the competition à customers switching banks<br/> <b>H5</b> Involuntary switching à customers switching banks<br/> <b>H6</b> distance à customers switching banks<br/> <b>H7</b> switching costs à customers switching banks<br/> <b>H8 -H15</b> Demographic Characteristic à customers switching banks</p> | <p><b>H1</b> Proportion of local calls to other networks à susceptibility to churn<br/> <b>H2</b> Proportion of STD calls to other networks à susceptibility to churn<br/> <b>H3</b> Usage of VASs à susceptibility to churn<br/> <b>H4</b> Overall usage RPM à susceptibility to churn</p> | <p><b>H1</b> Service bundling à Degree of contents usage<br/> <b>H2</b> Remaining contract months à Degree of contents usage<br/> <b>H3</b> Membership points à Degree of contents usage<br/> <b>H4</b> Total VOC à Degree of contents usage<br/> <b>H5</b> Membership years à Degree of contents usage<br/> <b>H6</b> Service bundling à Customer defection<br/> <b>H7</b> Remaining contract months à Customer defection<br/> <b>H8</b> Membership points à Customer defection<br/> <b>H9</b> Total VOC à Customer defection<br/> <b>H10</b> Membership years à Customer defection<br/> <b>H11</b> Channel views à Customer defection<br/> <b>H12</b> VOD views à Customer defection<br/> <b>H13</b> TVOD views à Customer defection<br/> <b>H14</b> PPV à Customer defection<br/> <b>H15</b> Monthly subscriptions à Customer defection</p> |
| <p><b>Results</b></p> <p>H1a, H1c, H2a, H3a, H4, H1c', H2a' and H3a' = Supported<br/> H1b, H2b, H3b, H3c, H1a', H1b', H2b', H3b' and H3c' = Rejected</p>   | <p>H1 to H15 = Supported</p>   | <p>H1 to H4 = Supported</p>   | <p>H1, H2, H3, H4, H5, H6, H9, H14 and H15 = Supported<br/> H7, H8, H10, H11, H12 and H13 = Rejected</p>   |

### Excellent service

An excellent service exceeds consumer expectations and satisfaction by providing high-quality services. Those are ability, attitude, appearance, attention, action, and accountability.<sup>20</sup> A customer-centric approach benefits both private and public companies. Using this mindset also helps providers win customers and save money. In addition, customers will stay loyal regardless of market choice if companies treat them well.

### Transparent subscriptions

In 2019, accountability was the key to sustaining a profitable company. Authenticity trumps traditional priorities like price and brand recognition.<sup>21</sup> In addition, transparency promotes trust, peace of mind and openness. Happy customers are more pleasant than dissatisfied customers, As such, individual customers feedback counts.

### Churn prediction techniques/framework

Companies can classify potential customers who leave the services using advanced machine learning (ML) technology. Then, using existing data, the company can identify potential churn customers. This knowledge would allow the company to target those customers and recover them. [Table 2](#) and [Table 3](#) summarise the most recent churn prediction and framework studies.

Machine learning can predict customer churn by identifying at-risk clients, pain points and interpreting data. [Table 3](#) identifies dependent and independent variables in prior research on customer churn prediction. The cause is an

**Table 4. Churn prediction techniques.**

| Algorithms                                 | Description   |
|--|---|
| Logistic Regression (LR)                   | Logistic regression is a practical regression analysis where the dependent variable is dichotomous; (binary). Logistic regression is used to characterise data and explain the relationship between one dependent binary variable and one or more nominal, ordinal, interval, or ratio-level independent variables. The odds ratio of multiple explanatory variables is calculated by logistic regression. Except that the response variable is binomial, the method is like multiple linear regressions. The result is the impact of each variable on the odds ratio of the event of interest. <sup>34</sup> |
| Linear Discriminant Analysis (LDA)         | Linear Discriminant Analysis, or LDA, is a technique used to minimise dimensionality. It is used as a pre-processing stage in applications for ML and pattern classification. The LDA aims to project the functions into a lower-dimensional space in a higher-dimensional space to avoid the curse of dimensionality and minimise energy and dimensional costs. <sup>35</sup>  |
| K-Nearest Neighbours Classifier (KNN)      | The Nearest Neighbour Classifier is a classification accomplished by defining the nearest neighbours as an example of a query and using those neighbours to evaluate the query's class. This classification approach is of particular interest since common run-time efficiency concerns are not the available computing resources these days. <sup>36</sup>  |
| Classification and Regression Trees (CART) | Classification and Regression Trees is a classification scheme that uses historical data to construct so-called decision trees. Decision trees can then be used in the form of new outcomes. First, the CART algorithm will search for all possible variables and all possible values to find the best partition—a query that divides the data into two parts with the highest homogeneity. Then, the process is repeated for each of the resulting data fragments. <sup>37</sup>   |
| Gaussian Naive Bayes (NB)                  | Gaussian Naive Bayes is a particular case of probabilistic networks that allows the treatment of continuous variables. It is a generalisation of Naive Bayes Networks. The Naive Bayes Classifiers are based on the Theorem of Bayes. One of the assumptions taken is the apparent presumption of freedom between functions. Furthermore, these classifiers assume that a particular function's value is unaffected by the value of any other feature. Therefore, naive Bayed Classifiers require a small amount of training data. <sup>38</sup>  |
| Support Vector Machine (SVM)               | Support vector machines (SVMs) are supervised learning methods known as regression, used for classification. Support vector machine (SVM) uses machine learning theory to classifier and regression prediction to maximise predictive accuracy while preventing overfitting the training. In general, SVMs may be thought of as systems that utilise functions in a high-dimensional feature space and are taught using an optimisation theory-based learning method that promotes statistical learning. <sup>39</sup>  |

independent variable, and the effect is a dependent variable. As such, the factors included in this study are customer churn, defection, demographics, and the voice of the customer (NPS rating).

### Traditional/statistical approaches

The traditional statistics approaches are used to solve problems involving less linear and repeatable data. They work well in environments with stable data and relationships. This is still widely used for medium- to long-term sales forecasting, where a reasonable forecast can be made with a few hundred or even fewer data points. Machine learning and statistics differ significantly. Machine learning models are created to make the most precise predictions possible. The purpose of statistical models is to make inferences about the relationships between variables.<sup>31</sup> Machine Learning is a data-driven algorithm that does not rely on rules-based programming. Statistical modelling is the use of mathematical equations to formalise relationships between variables.<sup>32</sup> This study applied machine learning and statistical approaches to analyse the potential churn and the mediation relationship between variables.

### Machine learning techniques in churn prediction

Machine Learning (ML) is a branch of artificial intelligence. ML uses existing algorithms and data sets to classify patterns.<sup>33</sup> This study adopts six widely used churn prediction techniques. All these algorithms were evaluated based on the performance accuracy when applied to the same data for a fair comparison. Table 4 describes the well-known churn prediction techniques.

## Research model

### Customer churn determinants

The following paragraphs explain the determinants of customer churn considered in this study. Figure 2 depicts four primary structures that may influence potential customer churn and the indirect effects of NPS feedback.

#### H1: Service Request

Companies recognise that poor customer service jeopardises customer relationships and revenue in the highly competitive telecommunications industry. Therefore, the degree to which telecommunications companies disconnect services indicates customer satisfaction and is directly proportional to customer turnover<sup>40</sup> (Table 5).

As competition grows and consumers place a higher value on service quality, service providers may find it increasingly difficult to succeed unless they pay greater attention to consumer reviews and concerns.<sup>41</sup>

#### H2: Helpdesk Staff

Several customers stop doing business with a company when they feel unappreciated, unable to get the information they want to speak to them or an unreasonably rude and unhelpful employee<sup>42</sup> (Table 6).

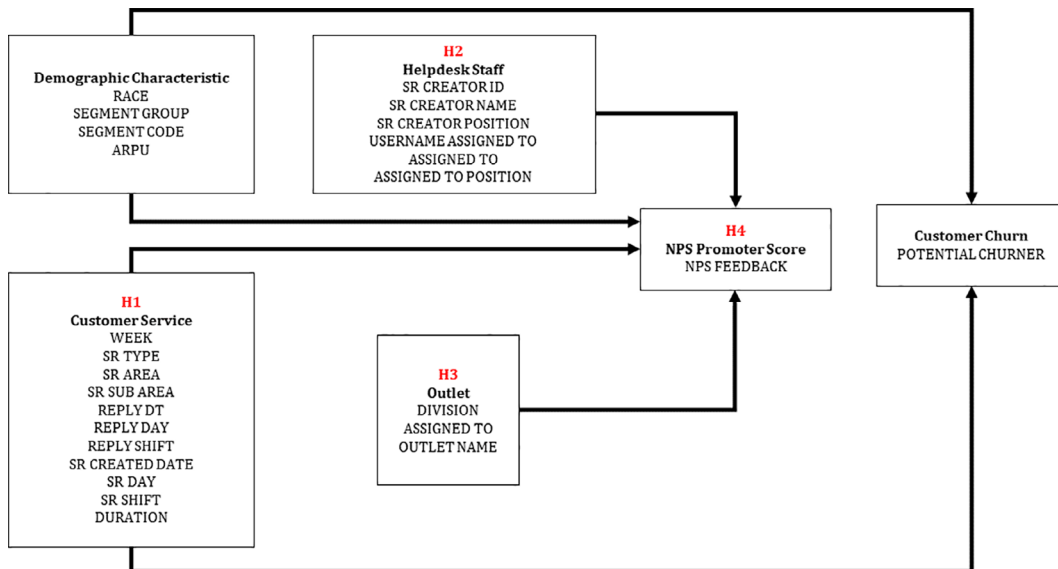


Figure 2. A conceptual model for the prediction of potential churning.



**Table 5. H1 Hypothesis.**

|     |  |
|-----|--|
| H1a | WEEK is positively associated with the customer churn probability            |
| H1b | SR TYPE is positively associated with the customer churn probability         |
| H1c | SR AREA is positively associated with the customer churn probability         |
| H1d | REPLY DT are positively associated with the customer churn probability       |
| H1e | REPLY DAY are positively associated with the customer churn probability      |
| H1f | REPLY SHIFT is positively associated with the customer churn probability     |
| H1g | SR CREATED DATE is positively associated with the customer churn probability |
| H1h | SR DAY are positively associated with the customer churn probability         |
| H1i | SR SHIFT is positively associated with the customer churn probability        |
| H1j | DURATION is positively associated with the customer churn probability        |

**Table 6. H2 Hypothesis.**

|     |  |
|-----|--|
| H2a | SR CREATOR ID is positively associated with the customer churn probability         |
| H2b | SR CREATOR NAME is positively associated with the customer churn probability       |
| H2c | SR CREATOR POSITION is positively associated with the customer churn probability   |
| H2d | USERNAME ASSIGNED TO are positively associated with the customer churn probability |
| H2e | ASSIGNED TO are positively associated with the customer churn probability          |
| H2f | ASSIGNED TO POSITION are positively associated with the customer churn probability |

Poor customer service, such as rude employees, delays in service, or incorrect details, can cause customer frustration and increase the churn rate.<sup>43</sup>

### H3: Outlet

Professionalism, friendliness, knowledge, communication, and sales skills are a few examples. Additionally, providers can reduce customer churn by adjusting service prices, policies, and branching<sup>44</sup> (Table 7).

An important management assumption is that employee attitudes and reactions to organisational changes are related to department performance.<sup>45</sup>

### H4: NPS score feedback

Collecting NPS surveys is a great way to get customer feedback and send them to the right team. For example, promoters should send a customer's name to the team for testimonials and case studies or sign up for a customer loyalty programme.<sup>46</sup> The survey divided over a thousand NPS feedback types into three categories: distractor, passive, and promoter (Table 8).

### Mediation effects of NPS feedback

A mediating variable links the independent and dependent variables. Its existence explains why the other two variables have a mediator relationship.<sup>47</sup> Some churn predictors may impact customer churn directly, indirectly, or both. This study

**Table 7. H3 Hypothesis.**

|     |  |
|-----|--|
| H3a | DIVISION ASSIGNED TO are positively associated with the customer churn probability |
| H3b | OUTLET NAME is positively associated with the customer churn probability           |

**Table 8. H4 Hypothesis.**

|    |   |
|----|---|
| H4 | A lower NPS feedback rating is considered more potential churning than a customer with a higher NPS feedback rating |
|----|---|

**Table 9. Mediation effects on NPS feedback.**

|      |   |
|------|---|
| H1a' | A NPS Feedback rating mediates the effect of request week on customer churn.                          |
| H1b' | A NPS Feedback rating mediates the effect of service request type on customer churn.                  |
| H1c' | A NPS Feedback rating mediates the effect of the service request area on customer churn               |
| H1d' | A NPS Feedback rating mediates the effect of respond date on customer churn                           |
| H1e' | A NPS Feedback rating mediates the effect of respond day on customer churn                            |
| H1f' | A NPS Feedback rating mediates the effect of respond day shift on customer churn                      |
| H1g' | A NPS Feedback rating mediates the effect of service request date on customer churn                   |
| H1h' | A NPS Feedback rating mediates the effect of service request day on customer churn                    |
| H1i' | A NPS Feedback rating mediates the effect of service request day shift on customer churn              |
| H1j' | A NPS Feedback rating mediates the effect of service duration on customer churn.                      |
| H2a' | A NPS Feedback rating mediates the effect of service request created staff ID on customer churn       |
| H2b' | A NPS Feedback rating mediates the effect of service request created staff name on customer churn     |
| H2c' | A NPS Feedback rating mediates the effect of service request created staff position on customer churn |
| H2d' | A NPS Feedback rating mediates the effect of assigned staff ID on customer churn                      |
| H2e' | A NPS Feedback rating mediates the effect of assigned staff name on customer churn                    |
| H2f' | A NPS Feedback rating mediates the effect of assigned staff position on customer churn                |
| H3a' | A NPS Feedback rating mediates the effect of assigned division on customer churn                      |
| H3b' | A NPS Feedback rating mediates the effect of the assigned outlet on customer churn                    |

defines partial defection as an NPS feedback rating from promoter to passive and total defection as passive to the distractor. Thus, it acts as a link between churn predictors and customer loss. Partial defection's mediation effects on churn determinants and total defection are investigated. Hence, an NPS feedback status is hypothesised to mediate the relationship (Table 9).

## Methods

This empirical study used 7776 random samples from a database of one of Malaysia's leading telecommunications service providers, spanning from two datasets ((MTD) September 2019 and (MTD) September 2020)) (Table 10). The primary key results from the discovery process can be seen in Table 11.

**Table 10. Dataset details.**

| No | Data                   | Data type | Original/adding data | Details   |
|----|------------------------|-----------|----------------------|---|
| 1  | SR NUMBER              | object    | Original data        | Service request tracking number   |
| 2  | WEEK                   | int64     | Original data        | Number of weeks (year)  |
| 3  | CUSTOMER NAME          | object    | Original data        | Customer/Business Name  |
| 4  | RACE                   | object    | Add data             | Customer Race   |
| 5  | REPLY DT               | object    | Original data        | Respond Date  |
| 6  | REPLY DAY              | object    | Add data             | Respond Day   |
| 7  | REPLY SHIFT            | object    | Add data             | Respond Day Shift   |
| 8  | S_NPS_FEEDBACK         | int64     | Original data        | NPS Feedback Rating (0-10)  |
| 9  | S_NPS_FEEDBACK_TYPE_FK | object    | Original data        | NPS Feedback Type (Promoter, Passive, Distractor)                         |
| 10 | NES RESPONSE           | object    | Original data        | NPS comment   |
| 11 | SEGMENT GROUP          | object    | Original data        | Customer segmentation group (consumer, SME's, government, and enterprise) |
| 12 | SEGMENT CODE           | object    | Original data        | Customer segmentation code  |
| 13 | ARPU                   | float64   | Add data             | The average revenue per user (customer)                                   |

**Table 10.** *Continued*

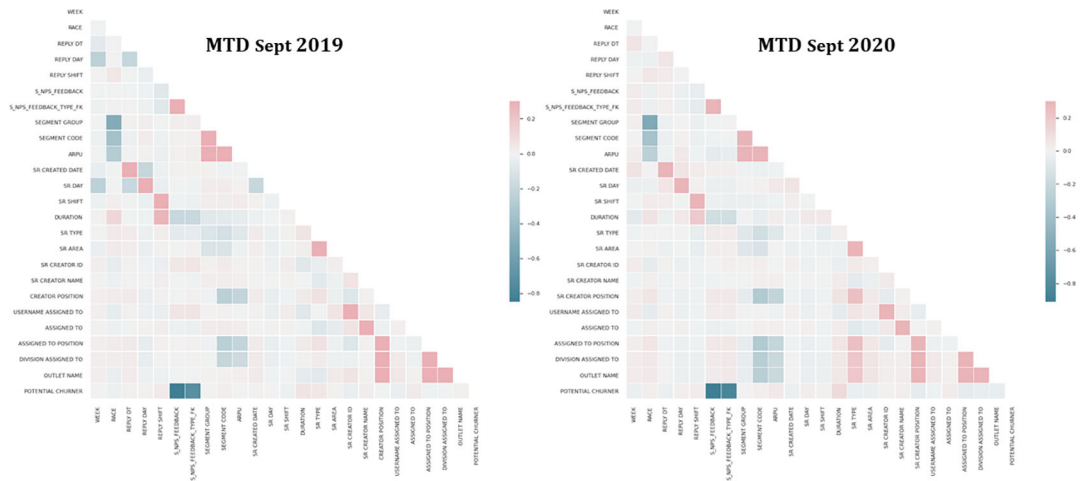
| No | Data                 | Data type | Original/adding data | Details                         |
|----|----------------------|-----------|----------------------|---------------------------------|
| 14 | SR CREATED DATE      | object    | Original data        | Service Request Date            |
| 15 | SR DAY               | object    | Add data             | Service Request Day             |
| 16 | SR SHIFT             | object    | Add data             | Service Request Day shift       |
| 17 | DURATION             | object    | Add data             | Respond Time duration           |
| 18 | SR TYPE              | object    | Original data        | Service Request Type            |
| 19 | SR AREA              | object    | Original data        | Service Request Area            |
| 20 | SR SUB AREA          | object    | Original data        | Service Request Sub Area        |
| 21 | SR CREATOR ID        | object    | Original data        | Helpdesk Staff Creator ID       |
| 22 | SR CREATOR NAME      | object    | Original data        | Helpdesk Staff Creator Name     |
| 23 | CREATOR POSITION     | object    | Original data        | Helpdesk Staff Creator Position |
| 24 | USERNAME ASSIGNED TO | object    | Original data        | Officer in Charge Username      |
| 25 | ASSIGNED TO          | object    | Original data        | Officer in Charge Name          |
| 26 | ASSIGNED TO POSITION | object    | Original data        | Officer in Charge Position      |
| 27 | DIVISION ASSIGNED TO | object    | Original data        | Assigned Division               |
| 28 | BUILDING ID          | object    | Original data        | Assigned Outlet ID              |
| 29 | OUTLET NAME          | object    | Original data        | Assigned Outlet Name            |
| 30 | ZONE                 | object    | Original data        | Assigned Outlet Zone            |
| 31 | STATE                | object    | Original data        | Assigned Outlet State           |
| 32 | SOURCE               | object    | Original data        | Service Request System          |
| 33 | POTENTIAL CHURNER    | object    | Add data             | Potential churner or not        |

**Table 11.** Key findings from the discovery process.

| MTD Sept 2019  | MTD Sept 2020   |
|--|---|
| <p><b>Total Dataset:</b> 4554<br/>Potential Churner: No: 4307 (94.58%) Yes: 247 (5.42%)</p>  | <p><b>Total Dataset:</b> 3222<br/>Potential Churner: No: 3101 (96.24%) Yes: 121 (3.76%)</p>   |
| <p><b>Customer segmentation</b><br/>Consumer: 4036, potential churner 219 = 5.4%<br/>Enterprise: 18, potential churner 1 = 5.6%<br/>Government: 24, potential churner 0 = 0%<br/>SME: 476, potential churner 27 = 5.7%<br/>From the total 4554 customers, Consumer segmentation is the highest segmentation interact with helpdesk staff and contribute to NPS feedback rating = 88.6%</p>   | <p><b>Customer segmentation</b><br/>Consumer: 2762, potential churner 102 = 3.7%<br/>Enterprise: 19, potential churner 1 = 5.6%<br/>Government: 20, potential churner 2 = 10%<br/>SME: 419, potential churner 16 = 3.8%<br/>From the total 3222 customers, Consumer segmentation is the highest segmentation interact with helpdesk staff and contribute to NPS feedback rating = 85.7%</p> |
| <p><b>NPS feedback rating</b><br/>Promoter: 3665, given a rating between 9-10<br/>Passive: 577, given a rating between 7-8<br/>Detractor: 312, given a rating between 0-6, potential churner 247 = 79.2% equivalence to given rating between 0-5<br/>9-10 (promoter) is the highest rating given by customer = 80.5% from overall 4554 customers</p>   | <p><b>NPS feedback rating</b><br/>Promoter: 2963, given a rating between 9-10<br/>Passive: 125, given a rating between 7-8<br/>Detractor: 134, given a rating between 0-6, potential churner 121 = 90.3% equivalence to given rating between 0-5<br/>9-10 (promoter) is the highest rating given by customer = 92.1% from overall 3222 customers</p>  |
| <p><b>Numerical value</b><br/>Week: the majority of 1271 customers communicates with the helpdesk staff in week 39, 3<sup>rd</sup> week of Sept 2019<br/>NPS feedback: the majority of 2436 customers were given a rating of 10.<br/>Average revenue per user (ARPU): 1275, R40 customers with APPU RM116.57 (Consumer) is the majority segmentation group interact with helpdesk staff.</p> | <p><b>Numerical value</b><br/>Week: the majority of 794 customers communicates with the helpdesk staff in week 39, 3<sup>rd</sup> week of Sept 2019<br/>NPS feedback: the majority of 2494 customers were given a rating of 10.<br/>Average revenue per user (ARPU): 1221, R40 customers with APPU RM116.40 (Consumer) is the majority segmentation group interact with helpdesk staff.</p> |

**Table 12. Variable selection and transformation.**

| No | Data                   | Data type |
|----|------------------------|-----------|
| 1  | WEEK                   | int64     |
| 2  | RACE                   | int64     |
| 3  | REPLY_DT               | int64     |
| 4  | REPLY_DAY              | int64     |
| 5  | REPLY_SHIFT            | int64     |
| 6  | S_NPS_FEEDBACK         | int64     |
| 7  | S_NPS_FEEDBACK_TYPE_FK | int64     |
| 8  | SEGMENT GROUP          | int64     |
| 9  | SEGMENT CODE           | int64     |
| 10 | ARPU                   | float64   |
| 11 | SR_CREATED_DATE        | int64     |
| 12 | SR_DAY                 | int64     |
| 13 | SR_SHIFT               | int64     |
| 14 | DURATION               | int64     |
| 15 | SR_TYPE                | int64     |
| 16 | SR_AREA                | int64     |
| 17 | SR_CREATOR_ID          | int64     |
| 18 | SR_CREATOR_NAME        | int64     |
| 19 | SR_CREATOR_POSITION    | int64     |
| 20 | USERNAME_ASSIGNED_TO   | int64     |
| 21 | ASSIGNED_TO            | int64     |
| 22 | ASSIGNED_TO_POSITION   | int64     |
| 23 | DIVISION_ASSIGNED_TO   | int64     |
| 24 | OUTLET_NAME            | int64     |
| 25 | POTENTIAL_CHURNER      | int64     |



**Figure 3. Correlation coefficient results for MTD Sept 2019 and MTD Sept 2020.**

### Variable selection and transformation

The dependent variables for potential churning are binary, with 1 representing “yes” and 0 representing “no”. In addition, a multinomial variable for each account indicates promoter, passive, and detractor NPS feedback. Thus, a positive correlation coefficient implies a direct relationship between the two variables. Conversely, inverse correlation occurs when one variable rises while the other falls.<sup>48</sup> Finally, after data pre-processing, eight irrelevant variables were dropped, and 25 variables were selected and converted numerically to avoid unstable coefficient estimates and difficult model interpretation (Table 12) (Underlying data).<sup>49</sup>

The study found high correlations between variables. Figure 3 shows the most positive correlation between helpdesk staff and assigned officer in charge ( $r = 0.98$  &  $0.96$ ) and the most negative correlation between NPS feedback and potential churning ( $r = -0.85$  &  $-0.91$ ). The potential churning is found to be negatively related to NPS Feedback. Customers with lower NPS ratings are more likely to churn than those with higher ratings.

## Results

### Machine learning algorithm

This study tested a harness to use 10-fold cross-validation, builds multiple models to predict measurements, and selects the best model. As a result, CART has the highest estimated accuracy score of 0.98 or 98% (Table 13).

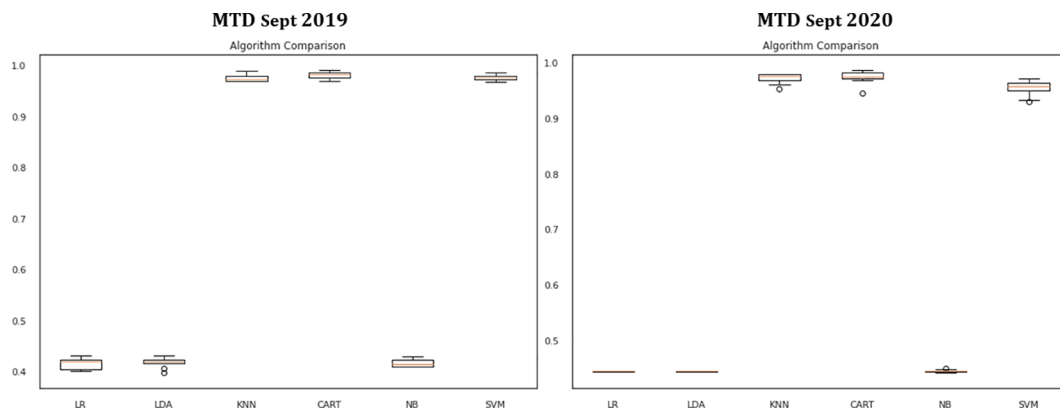
In Figure 4, the box and whisker plots at the top of the range, with CART, SVM, and KNN evaluations achieving 100% accuracy and NB, LDA, and LR evaluations falling into the low 41% accuracy range.

### Mediation analysis results

Table 14 shows the findings of the mediation effects, with statistical significance presented as a p-value less than 0.05. According to the results of this study, the NPS feedback rating appears to be a partial mediator between some churn determinants and customer churn. NPS Feedback is found to be a significant mediator of several churn determinants. The

**Table 13. Machine learning algorithms comparison results.**

| Algorithms Name                            | MTD Sept 2019 |      |                | MTD Sept 2020 |      |                |
|--|---------------|------|----------------|---------------|------|----------------|
|  | Mean          | Std  | Accuracy score | Mean          | Std  | Accuracy score |
| Logistic Regression (LR)                   | 0.42          | 0.01 | 41%            | 0.44          | 0.01 | 45%            |
| Linear Discriminant Analysis (LDA)         | 0.41          | 0.02 | 42%            | 0.47          | 0.02 | 45%            |
| K-Nearest Neighbours Classifier (KNN)      | 0.98          | 0.01 | 98%            | 0.98          | 0.01 | 97%            |
| Classification and Regression Trees (CART) | 0.98          | 0.01 | 98%            | 0.98          | 0.01 | 98%            |
| Gaussian Naive Bayes (NB)                  | 0.41          | 0.01 | 41%            | 0.44          | 0.02 | 44%            |
| Support Vector Machine (SVM)               | 0.98          | 0.01 | 98%            | 0.96          | 0.01 | 98%            |



**Figure 4. Comparing machine learning algorithms for MTD Sept 2019 and MTD Sept 2020.**

**Table 14. Mediation analysis results.**

| <b>X: Independent variable</b> |                              |                              |  |   |
|--------------------------------|------------------------------|------------------------------|--|---|
| <b>M: NPS feedback</b>         |                              |                              |  |   |
| <b>Y: Potential churner</b>    |                              |                              |  |   |
| <b>X</b>                       | <b>X and M<br/>(p-value)</b> | <b>X and Y<br/>(p-value)</b> | <b>X, M and Y<br/>(p-value) sobel test</b> | <b>Significant relationship<br/>X and Y via M</b> |
| WEEK                           | 0.5365                       | 0.7852                       | 0.5365                                     | No  |
| SR TYPE*                       | 0.0001                       | 0.0024                       | 0.0001                                     | Yes   |
| SR AREA                        | 0.6196                       | 0.7088                       | 0.6195                                     | No  |
| REPLY DT                       | 0.5634                       | 0.8215                       | 0.5634                                     | No  |
| REPLY DAY                      | 0.4331                       | 0.4755                       | 0.4331                                     | No  |
| REPLY SHIFT*                   | 0.0001                       | 0.0138                       | 0.0001                                     | Yes   |
| SR CREATED DATE                | 0.7494                       | 0.8044                       | 0.7494                                     | No  |
| SR DAY                         | 0.7311                       | 0.7859                       | 0.7311                                     | No  |
| SR SHIFT                       | 0.4008                       | 0.2146                       | 0.4008                                     | No  |
| DURATION*                      | 0.0001                       | 0.0001                       | 0.0001                                     | Yes   |
| SR CREATOR ID*                 | 0.0009                       | 0.0653                       | 0.0009                                     | Yes   |
| SR CREATOR NAME                | 0.5596                       | 0.1048                       | 0.5596                                     | No  |
| SR CREATOR POSITION            | 0.7382                       | 0.7535                       | 0.7382                                     | No  |
| USERNAME ASSIGNED TO*          | 0.0002                       | 0.0292                       | 0.0002                                     | Yes   |
| ASSIGNED TO                    | 0.1727                       | 0.0206                       | 0.1727                                     | No  |
| ASSIGNED TO POSITION           | 0.9376                       | 0.4814                       | 0.9376                                     | No  |
| DIVISION ASSIGNED TO           | 0.9622                       | 0.6390                       | 0.9622                                     | No  |
| OUTLET NAME                    | 0.6197                       | 0.5280                       | 0.6197                                     | No  |

\*p-value (<.05).

NPS feedback rating change partially mediates the following variables' effects on the probability of customer churn, Duration, Reply Shift, Service Request Type, Helpdesk Staff ID, and Assigned Officer to handle the task have a significant relationship with potential churn customer.

#### Hypothesis test results: Customer churn determinants

H1b reveals that SR TYPE (Service Request Type) has a significant impact on the probability of churn (Table 15). This finding is supported by Solution Partner (2019) that customers' requests should be centralised to avoid multiple ticket opening sources. One-stop-centre to answer requests, provide helpful information, and engage with customers until the problem is solved.<sup>50</sup>

H1f reveals that REPLY SHIFT (Respond Day Shift) significantly impacts the probability of churn. It is consistent with past research; the speed of customer service responses is also important. Thus, the solution can influence employee and customer engagement when seeking solutions.<sup>51</sup>

H1j reveals that DURATION (Respond Time Duration) has a significant impact on the probability of churn. This result supports Scout (2020) that time is an important factor in determining customer service quality. According to Forrester Research, 77% of customers believe that respecting their time is the most important online customer service.<sup>52</sup>

H2d reveals that USERNAME ASSIGNED (Officer in Charge Username) significantly impacts the probability of churn. However, H2f does not. As a result, help desk staff must assign problems to experts based on case-by-case categories. Adebisi *et al.* (2016) found that failing to respond to customer complaints or provide solutions may result in poor service delivery and contract termination.<sup>53</sup>

**Table 15. Summary of hypothesis results.**

| Hypothesis | Description  | Decision  |
|------------|--|-----------|
| H1a        | WEEK is positively associated with the customer churn probability  | Rejected  |
| H1b        | SR TYPE is positively associated with the customer churn probability   | Supported |
| H1c        | SR AREA is positively associated with the customer churn probability   | Rejected  |
| H1d        | REPLY DT are positively associated with the customer churn probability   | Rejected  |
| H1e        | REPLY DAY are positively associated with the customer churn probability  | Rejected  |
| H1f        | REPLY SHIFT is positively associated with the customer churn probability   | Supported |
| H1g        | SR CREATED DATE is positively associated with the customer churn probability                                       | Rejected  |
| H1h        | SR DAY are positively associated with the customer churn probability   | Rejected  |
| H1i        | SR SHIFT is positively associated with the customer churn probability  | Rejected  |
| H1j        | DURATION is positively associated with the customer churn probability  | Supported |
| H2a        | SR CREATOR ID is positively associated with the customer churn probability   | Supported |
| H2b        | SR CREATOR NAME is positively associated with the customer churn probability                                       | Rejected  |
| H2c        | SR CREATOR POSITION is positively associated with the customer churn probability                                   | Rejected  |
| H2d        | USERNAME ASSIGNED TO are positively associated with the customer churn probability                                 | Supported |
| H2e        | ASSIGNED TO are positively associated with the customer churn probability  | Rejected  |
| H2f        | ASSIGNED TO POSITION are positively associated with the customer churn probability                                 | Rejected  |
| H3a        | DIVISION ASSIGNED TO are positively associated with the customer churn probability                                 | Rejected  |
| H3b        | OUTLET NAME is positively associated with the customer churn probability   | Rejected  |
| H4         | A lower NPS feedback rating is considered more potential churner than a customer with a higher NPS feedback rating | Supported |

## Discussion

This study found that customer satisfaction with helpdesk service affects NPS scores. Each rating meets customers' expectations. Understanding the provider's potential churner will also reveal how the company operates, whether it provides a high-quality product with excellent customer service or needs to improve significantly to compete.

It is important to acknowledge customers' contributions to the product or service's value. Responding to complaints is a good start, but if the provider wants to stay in business, they will need to do more. Providers will be delighted if they meet all customers' needs while providing the highest quality services.<sup>54</sup>

From 7776 records, 5% of customers with NPS ratings below 6 are potential churners. The Customer Relationship Management (CRM) team will find potential churners with the initiative and techniques to retain customers using the same framework. Pope (2020) research the customers' lifetime value entirely depends on how hard the businesses work to maintain them. Providing a personalised customer experience will keep them coming back. It will also turn ardent supporters into online, social media, and in-person brand advocates. But building the brand momentum through happy customers takes time and effort. A good product is not enough. When consumers make purchases, they anticipate an experience. Numerous businesses use retention marketing to ensure brand consistency. The third and final phase, retention marketing, is the most critical. Providers must focus on customer relationships.<sup>55</sup>

## Conclusions

This research used data from September 2019 and September 2020 to predict churn potential among Malaysian telecommunications customers. According to the results, the immediate helpdesk response can ensure that customers' needs are met and act as mediators in determining an employee's ability to satisfy customers. This study evaluated six machine learning algorithms, with CART having the most accurate performance (98%). The NPS feedback rating partially mediates customer churn. The proposed framework will help providers accurately predict potential churn customers and help CRM teams offer targeted churning customers win-back programmes. For better findings and analysis, more research should be done on NPS rating and provided customer feedback.

## Data availability

### Underlying data

Zenodo: Customer Churn Prediction for telecommunication Industry: A Malaysian Case Study.

DOI: <https://doi.org/10.5281/zenodo.5758742>.<sup>49</sup>

This project contains the following underlying data:

- Dataset MTD Sept 2019.csv (The file contains Net Promoter Score (NPS) in Month to Date (MTD) September 2019 of a telecommunication company. 25 variables were used to determine potential churn customer).
- Dataset MTD Sept 2020.csv (The file contains Net Promoter Score (NPS) in Month to Date (MTD) September 2020 of a telecommunication company. 25 variables were used to determine potential churn customer).

Data are available under the terms of the [Creative Commons Attribution 4.0 International license](#) (CC-BY 4.0).

## Author contributions

Nurulhuda, M., Lew, S. L., & Siti, F. A. R. comprehended the idea and contributed to the research article. All authors contributed to the writing, editing, and consent of the final manuscript.

## Acknowledgements

First and foremost, I must thank Multimedia University, for the sponsorship towards my Master of Science (Information Technology). Lastly, special thanks to all my family members and friends for their continuous support and understanding.

## References

1. Tamaddoni Jahromi A, Stakhovych S, Ewing M: **Managing B2B customer churn, retention and profitability**. *Ind. Mark. Manag.* 2014; **43**(7): 1258–1268. [Publisher Full Text](#)
2. Angelova B, Zekiri J: **Measuring Customer Satisfaction with Service Quality Using American Customer Satisfaction Model (ACSI Model)**. *International Journal of Academic Research in Business and Social Sciences*. 2011; **1**(3): 27. [Publisher Full Text](#)
3. Hejazinia R, Kazem M: **Prioritising factors influencing customer churn**. *Interdisciplinary Journal of Contemporary Research in Business*. 2014; **5**(12): 1–10. [Reference Source](#)
4. Kumar S, Chandrakala D: **A Survey on Customer Churn Prediction using Machine Learning Techniques**. *Int. J. Comput. Appl.* 2016; **154**(10): 13–16. [Publisher Full Text](#)
5. Ahmad AK, Jafar A, Aljoumaa K: **Customer churn prediction in telecom using machine learning in big data platform**. *J. Big Data*. 2019; **6**(1): 6–28. [Publisher Full Text](#)
6. Ahn J-H, Han S-P, Lee Y-S: **Customer churn analysis: Churn determinants and mediation effects of partial defection in the Korean mobile telecommunications service industry**. *Telecommun. Policy*. 2006; **30**(10–11): 552–568. [Publisher Full Text](#)
7. Global Legal Group: **Telecoms, Media & Internet 2021 | Malaysia | ICLG. International Comparative Legal Guides International Business Reports**. 2020, 11th December. [Reference Source](#)
8. Idem Est Advisory & Research: **Latest Malaysia Telecoms Industry Report – 2020-2025**. Idem Est Research & Advisory. 2020, 31st August. [Reference Source](#)
9. Markets RA: **Malaysia Telecoms Industry Report 2020-2025 - Market Sizing, Insights, Key Telecom Trends, 5G, IoT, and More**. *GlobeNewswire Newsroom*. 2020, 21st February. [Reference Source](#)
10. Churn Rate: **Investopedia**. 2021, 1st March. [Reference Source](#)
11. Reading Churn Rates: **Investopedia**. 2020, 25th May. [Reference Source](#)
12. **The Influence of Perceived Organisational Justice on Customer's Trust: An Overview of Public Higher Educational Students**. *Int. J. Bus. Manag.* 2019; **3**(3): 01–08. [Publisher Full Text](#)
13. 83% of consumers in Malaysia may switch telco provider: Frost: **Digital News Asia**. 2015, 7th April. [Reference Source](#)
14. What is Net Promoter Score? (Updated 2020): **Qualtrics AU**. 2020, 12th October. [Reference Source](#)
15. Gadkari D: *Factors Influencing the Net Promoter Score*. UPPSALA UNIVERSITET; 2018, June. [Reference Source](#)
16. Keiningham TL, Aksoy L, Cooil B, et al.: **A holistic examination of Net Promoter**. *J. Database Mark. Cust. Strategy Manag.* 2008; **15**(2): 79–90. [Publisher Full Text](#)
17. What is Net Promoter Score (NPS)? 2021 Guide & Definition: **Hotjar**. 2021. [Reference Source](#)
18. Nunkoo R, Teeroovengadam V, Ringle CM, et al.: **Service quality and customer satisfaction: The moderating effects of hotel star rating**. *Int. J. Hosp. Manag.* 2020; **91**: 102414. [Publisher Full Text](#)
19. Rajeev M: **What is Customer Satisfaction?** *Freshcaller Blog*. 2020, 18th May. [Reference Source](#)
20. Calvert GA, Pathak A, Ching LEA, et al.: **Providing Excellent Consumer Service Is Therapeutic: Insights from an Implicit Association Neuromarketing Study**. *Behav. Sci.* 2019; **9**(10): 109. [PubMed Abstract](#) | [Publisher Full Text](#)
21. Kundeliene K, Leitoniene S: **Business Information Transparency: Causes and Evaluation Possibilities**. *Procedia. Soc. Behav. Sci.* 2015; **213**: 340–344. [Publisher Full Text](#)
22. Ahmad AK, Jafar A, Aljoumaa K: **Customer churn prediction in telecom using machine learning in big data platform**. *Journal of*



- Big Data*. 2019; **6**(1): 1–5.  
[Publisher Full Text](#)
23. Höppner S, Stripling E, Baesens B, et al.: **Profit-driven decision trees for churn prediction**. *Eur. J. Oper. Res.* 2020; **284**(3): 920–933.  
[Publisher Full Text](#)
  24. Yang L: **Prediction Modeling and Analysis for Telecom Customer Churn in Two Months**. *ArXiv.Org*. 2019, 1st November.  
[Reference Source](#)
  25. Ahmed AAQ, Maheswari D: **Churn prediction on huge telecom data using hybrid firefly-based classification**. *Egypt. Inform. J.* 2017; **18**(3): 215–220.  
[Publisher Full Text](#)
  26. Eria K, Marikannan BP: **Systematic Review of Customer Churn Prediction in the Telecom Sector**. *ResearchGate*. 2018, September 27.  
[Reference Source](#)
  27. Ahn J-H, Han S-P, Lee Y-S: **Customer churn analysis: Churn determinants and mediation effects of partial defection in the Korean mobile telecommunications service industry**. *Telecommun. Policy*. 2006; **30**(10–11): 552–568.  
[Publisher Full Text](#)
  28. Clemes MD, Gan C, Zhang D: **Customer switching behaviour in the Chinese retail banking industry**. *Int. J. Bank Mark.* 2010; **28**(7): 519–546.  
[Publisher Full Text](#)
  29. Geetha M, Abitha Kumari J: **Analysis of churn behavior of consumers in Indian telecom sector**. *J. Indian Bus. Res.* 2012; **4**(1): 24–35.  
[Publisher Full Text](#)
  30. Kim M-J, Kim J, Park S-Y: **Understanding IPTV churning behaviors: focus on users in South Korea**. *Asia Pacific Journal of Innovation and Entrepreneurship*. 2017; **11**(2): 190–213.  
[Publisher Full Text](#)
  31. Stewart MPR: **The Actual Difference Between Statistics and Machine Learning**. *Medium*. 2020, July 30.  
[Reference Source](#)
  32. Srivastava T: **Machine Learning vs. Statistical Modeling**. *Analytics Vidhya*. 2020, June 26.  
[Reference Source](#)
  33. Analytics P: **What Is Machine Learning: Definition, Types, Applications and Examples**. *Potentia Analytics*. 2019, December 19.  
[Reference Source](#)
  34. Bertocelli CM, Altamura P, Vieira ER, et al.: **PredictMed: A logistic regression-based model to predict health conditions in cerebral palsy**. *Health Informatics J.* 2020; **26**(3): 2105–2118.  
[Publisher Full Text](#)
  35. Tharwat A, Gaber T, Ibrahim A, et al.: **Linear discriminant analysis: A detailed tutorial**. *AI Commun.* 2017; **30**(2): 169–190.  
[Publisher Full Text](#)
  36. Cunningham P, Delany SJ: *K-Nearest Neighbour Classifiers*. 2nd ed. Machine Learning; 2020; **2**: 1–15. (with Python examples).  
[Reference Source](#)
  37. Fearn T: **Classification and Regression Trees (CART)**. *NIR News*. 2006; **17**(6): 13–14.  
[Publisher Full Text](#)
  38. Moraes RM, Machado LS: **Gaussian Naive Bayes for Online Training Assessment in Virtual Reality-Based Simulators**. *Mathware & Soft Computing*. 2009; **16**(2009): 123–132.  
[Reference Source](#)
  39. Jakkula V: *Tutorial on Support Vector Machine (SVM)*. School of EECS, Washington State University. 2006; 1–13.  
[Reference Source](#)
  40. Adsit DJ, London M, Crom S, et al.: **Relationships between employee attitudes, customer satisfaction and departmental performance**. *J. Manag. Dev.* 1996; **15**(1): 62–75.  
[Publisher Full Text](#)
  41. Kierczak L: **Customer Satisfaction: Why It's Still Important in 2021**. *Survicate*. 2021, January 21.  
[Reference Source](#)
  42. Council YE: **The Customer Is Not Always Right: Here Are Five Reasons Why**. *Forbes*. 2019, May 23.  
[Reference Source](#)
  43. MacDonald S: **Customer Complaints: Why Angry Customers Are Good for Business**. *SuperOffice*. 2021, May 4.  
[Reference Source](#)
  44. Hammah CA: **A customer retention strategy for Phoenix Insurance Company**. *Ashesi Institutional Repository*. 2021, July 20.  
[Reference Source](#)
  45. Lin CY, Huang CK: **Employee turnover intentions and job performance from a planned change: the effects of an organisational learning culture and job satisfaction**. *Int. J. Manpow.* 2020; **42**(3): 409–423.  
[Publisher Full Text](#)
  46. Baehre S, O'Dwyer M, O'Malley L, et al.: **The use of Net Promoter Score (NPS) to predict sales growth: insights from an empirical investigation**. *J. Acad. Market. Sci.* 2021.  
[Publisher Full Text](#)
  47. Walters GD: **Why are mediation effects so small?**. *Int. J. Soc. Res. Methodol.* 2018; **22**(2): 219–232.  
[Publisher Full Text](#)
  48. Schober P, Boer C, Schwarte LA: **Correlation Coefficients**. *Anesth. Analg.* 2018; **126**(5): 1763–1768.  
[Publisher Full Text](#)
  49. Mustafa N: **Customer Churn Prediction for Telecommunication Industry: A Malaysian Case Study [Data set]**. *Zenodo*. 2021;  
[Publisher Full Text](#)
  50. Solution Partner: **Best practices for Service Request Management**. *E-Core Consulting*. 2019, 20th August.  
[Reference Source](#)
  51. Tri Marwanto S, Komaladewi R: **How to Restrain Customer Churn in Telecommunication Providers: Study in West Java Indonesia**. *Review of Integrative Business and Economics Research*. 2017; **6**(1): 51–59.  
[Reference Source](#)
  52. Scout H: **Email Response Times: Benchmarks and Tips for Support**. *HelpScout*. 2020, June 9.  
[Reference Source](#)
  53. Adebisi SO, Oyatoye EO, Amole BB: **Relevant Drivers for Customers' Churn and Retention Decision in the Nigerian Mobile Telecommunication Industry**. *Journal of Competitiveness*. 2016; **6**(3): 52–67.  
[Publisher Full Text](#)
  54. Suchánek P, Králová M: **Customer satisfaction, loyalty, knowledge and competitiveness in the food industry**. *Economic Research-Ekonomska Istraživanja*. 2019; **32**(1): 1237–1255.  
[Publisher Full Text](#)
  55. Pope L: **How to Prevent Customer Churn with Retention Marketing**. *G2*. 2020, 27th August.  
[Reference Source](#)

# Open Peer Review

Current Peer Review Status:  

---

Version 1

Reviewer Report 28 April 2022

<https://doi.org/10.5256/f1000research.77258.r129495>

© 2022 Amalarethinam D. This is an open access peer review report distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.



**D. I. George Amalarethinam** 

Department of Computer Science, Jamal Mohamed College, Tiruchirappalli, Tamil Nadu, India

The salient features of gait are analysed to predict churn potential using the data from one of the Malaysia's standard telecommunications service providers with the specific period from September 2019 to September 2020.

- Additional explanation to be provided towards data called for in tables.
- Though the paper is technically sound, the limitations or negative aspects of the proposed methodology needs to be included.
- The recent year references may be provided towards strengthening of the paper.

**Is the work clearly and accurately presented and does it cite the current literature?**

Yes

**Is the study design appropriate and is the work technically sound?**

Yes

**Are sufficient details of methods and analysis provided to allow replication by others?**

Partly

**If applicable, is the statistical analysis and its interpretation appropriate?**

Not applicable

**Are all the source data underlying the results available to ensure full reproducibility?**

Yes

**Are the conclusions drawn adequately supported by the results?**

Partly

**Competing Interests:** No competing interests were disclosed.

**Reviewer Expertise:** Cloud Scheduling & Security, Data Mining, Network Security

**I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard.**

Reviewer Report 04 January 2022

<https://doi.org/10.5256/f1000research.77258.r109266>

© 2022 Supriyanto C. This is an open access peer review report distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.



**Catur Supriyanto** 

Faculty of Computer Science, Universitas Dian Nuswantoro, Semarang, Indonesia

The research paper title on “Customer churn prediction for telecommunication industry: A Malaysian Case Study” is well organized and the originality of the paper is good. The main objective of the paper is to compare some machine learning (ML) algorithms for customer churn prediction in telecommunication industry.

In experiment the authors analyzed that the data were collected from one of the Malaysia's leading telecommunications service providers. The authors hypothesized that changes in Net Promoter Score (NPS) can indicate whether churn determinants directly or indirectly influence churn. Hence, they decided to execute many algorithms such as Logistic Regression, K-Nearest Neighbor (KNN), Support Vector Machine (SVM), Linear Discriminant Analysis (LDA), Classification and Regression Trees (CART), Gaussian Naive Bayes (NB) algorithms.

The authors concluded that the CART outperforms with an accuracy of 98% than other ML algorithms. The structure of the paper and result findings are good. Finally, the outcome of the paper is good and suggested to other businesses to measure potential customer churn using NPS scores to gather customer feedback in near future.

Hence, I have recommended this paper to be indexed.

**Is the work clearly and accurately presented and does it cite the current literature?**

Yes

**Is the study design appropriate and is the work technically sound?**

Yes

**Are sufficient details of methods and analysis provided to allow replication by others?**

Yes

**If applicable, is the statistical analysis and its interpretation appropriate?**

Yes

**Are all the source data underlying the results available to ensure full reproducibility?**

Yes

**Are the conclusions drawn adequately supported by the results?**

Yes

**Competing Interests:** No competing interests were disclosed.

**Reviewer Expertise:** Machine learning

**I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard.**

---

The benefits of publishing with F1000Research:

- Your article is published within days, with no editorial bias
- You can publish traditional articles, null/negative results, case reports, data notes and more
- The peer review process is transparent and collaborative
- Your article is indexed in PubMed after passing peer review
- Dedicated customer support at every stage

For pre-submission enquiries, contact [research@f1000.com](mailto:research@f1000.com)

**F1000Research**