A Survey on Churn Analysis in Various Business Domains

JAEHYUN AHN¹, JUNSIK HWANG¹, DOYOUNG KIM¹, HYUKGEUN CHOI¹, AND SHINJIN KANG², (Member, IEEE)
¹Intelligence Laboratories, Nexon Korea, Seongnam 13487, Republic of Korea
²School of Games, Hongik University, Sejong Campus, Sejong 30016, Republic of Korea

Corresponding author: Shinjin Kang (directx@hongik.ac.kr)

This work was supported in part by the National Research Foundation of Korea, Korean government (MSIT), under Grant NRF-2019R1A2C1002525, and in part by Nexon Korea.

ABSTRACT In this paper, we present churn prediction techniques that have been released so far. Churn prediction is used in the fields of Internet services, games, insurance, and management. However, since it has been used intensively to increase the predictability of various industry/academic fields, there is a big difference in its definition and utilization. In this paper, we collected the definitions of churn used in the fields of business administration, marketing, IT, telecommunications, newspapers, insurance and psychology, and described their differences. Based on this, we classified and explained churn loss, feature engineering, and prediction models. Our study can be used to select the definition of churn and its associated models suitable for the service field that researchers are most interested in by integrating fragmented churn studies in industry/academic fields.

INDEX TERMS Churn analysis, churn prediction, machine learning, log analysis, data mining, customer retention, customer relationship management.

I. INTRODUCTION

The term customer churn is commonly used to describe the propensity of customers who cease doing businesses with a company in a given time or contract [1]. Traditionally, studies on customer churn started from Customer Relation Management (CRM) [2]. It is crucial to prevent customer churn when operating services. In the past, the efficiency of customer acquisition relative to the number of churns was good. However, as the market saturated because of the globalization of services and fierce competition, customer acquisition costs rose rapidly [3], [4].

Reinartz, Werner, Jacquelyn S. Thomas, and Viswanathan Kumar. (2005) have shown that, for long-term business operations, putting efforts to increase the retention rate of all customers in terms of CRM is less efficient than putting efforts on a small number of targeted customer acquisition activities [22]. Similarly, Sasser, W. Earl. (1990) have suggested that retained customers generally return higher margins than randomly targeting new customers [23]. Additionally, Mozer, Michael C., et al. (2000) have proposed that, in terms of net return on investment, marketing campaigns for retaining existing customers are more efficient than putting efforts to attract new customers [16]. Reichheld et al. (1996) have shown that a 5 percent increase in customer retention rate achieved 35 percent and 95 percent increases in net present value of customers for a software company and an advertising agency, respectively [29]. As such, churn prediction can be used as a method to increase the retention rate of loyal customers and ultimately increase the value of the company.

Studies on customer churn have been proposed in various service fields. These studies on the churn analysis attempted to identify or predict in advance the likelihood that customers will churn using various indicators. The customer churn rate [5] is a typical customer churn analysis indicator. This refers to the ratio of subscribers who cancel a service to the total number of subscribers during a specific period [5], [32], [41]. The churn rate is the most widely used indicator for calculating the service retention period of subscribers in most service fields. Because of its importance and intuition, churn has been introduced in various service fields and developed to suit the characteristics of each field. Consequently, the research on the analysis of customer churn was...
fragmented according to each research field, thus the measurement criteria are all different. Currently, this is causing many problems. In the industry, communication costs arising from different churn criteria between service personnel in the process of fusing heterogeneous services (e.g., vehicle sharing service/insurance, online music service/department store) have been sharply increasing. Furthermore, since research on churn is simultaneously associated with two fields of engineering and business administration, it is not easy for researchers to describe two separate specialized fields on a single paper or to understand them.

In the past, customer churn of early days was used to define the customer’s status in the CRM. The CRM is a business management method that first emerged as a way of increasing the efficiency in areas of retail, marketing, sales, customer service, and supply-chain, and increasing efficiency and the customer value functions of the organization [2]. Since then, in the architectural point of view, the CRM has evolved and become divided into operational CRM and analytical CRM. The analytical CRM is focused on developing databases and resources containing customer characteristics and attitudes [24], [25], [30]. The analytical CRM has been initially used for creating appropriate marketing strategies using customer status and customer behavior data, and particularly, it has been used to fulfill the individual and unique needs of customers [26]. From this point on, IT and knowledge management related technologies have been utilized, and companies started applying dedicated technologies for acquisition, retention, churn, and selection of customers [27], and ever since the technologies of IT field became implemented in the CRM, various companies began to use such technologies in business areas including data warehouse, website, telecommunication, and banking [28]. As described earlier, with studies on CRM claiming that increasing the retention rate of small number of existing customers is more efficient than acquiring new customers, churn analysis has become one of the important personalized customer management techniques [20], [28], [59]. There were survey papers that collected and summarized churn analysis techniques in the telecommunications field [6]–[9]. However, these studies are limited to the telecommunications field, and the log data used for the churn analysis do not include time series features, retention and survival, and KPI (Key Performance Indicator) features. There were also papers applied to services using various deep learning model-based churn analysis techniques in terms of computer science [10], [11]. However, these studies are limited to the deep learning algorithm, and lack underlying models and parameter description. There are also a few survey papers on churn, yet they do not cover the latest deep learning techniques but cover only churn in specific industrial fields [12], [13]. The trend of building churn prediction models is changing, and performance is rapidly improving. However, because of fragmented previous studies, there are many difficulties for researchers to launch new research on churn. In order to address these issues, this survey paper describes the differences in the definition of churn prediction algorithms in the fields of business administration, marketing, IT, telecommunications, newspaper publishing, insurance, and psychology, and compares differences in churn loss and feature engineering. In addition, we classify and explain the cases of churn prediction models based on this. Our study provides classification information for more detailed technologies on churn in a wider range than previous survey papers. Our research can reduce confusion about the churn criteria that are being fragmented and utilized across multiple industry/academic fields, and can be of a practical help in applying them to prediction models. In particular, this paper presents a deep learning model among machine learning techniques designed to solve non-contractual customer churns, which have recently appeared with the advancement of industries. The structure of this paper is as follows. Chapter II introduces typical definitions for churn in each business domain and their differences. Chapter III presents churn application cases in various business domains Chapters IV and V introduce losses and features used in Churn, respectively. Chapter VI introduces typical churn-based prediction models by classifying them according to each algorithm, and presents which algorithm is mainly used in each industry.

II. DEFINITION OF CHURN

Churn has been defined in various ways in multiple industries. In this chapter, we describe two typical types. As seen in Table 1, typical papers with different criteria for defining churn are summarized. In general, the dictionary definition of churn is known as the prolonged period of inactivity [14]. However, the criteria for ‘inactivity’ and ‘prolonged’ are different according to each research field. Such inconsistency is frequently found due to more services of modern days adopting loose subscription terms because of competition. In the past, customer churn had occurred explicitly through contractual cancellations, however, in the modern services including Internet and retail services, frequent customer churns occur due to the low customers’ investment costs [19]–[21], [96]. These non-contractual customer churns occur due to low switching cost for changing the service [31]. Thus, we can divide the criteria of churn into contractual churn and non-contractual churn. Descriptions of each churn is as follows.

The first criterion is contractual churn. Contractual churn refers to churn that a customer does not extend the contract even when the contract renewal date is reached [15], [16]. This churn means that a customer loses interest in the relevant service area and changes his/her position to a state where re-entry is no longer possible. It is usually present in churn problems occurring when customers close their banking accounts or when switching their carrier operator from one service to another. In addition, contractual churn is frequently found in a flat-rate service such as music and movie streaming services.

The second criterion is non-contractual churn. In general, in a non-contractual situation, customers can leave the service/contract without time constraints. In the service operating perspective, a criterion for churn is first constructed, then a customer that meets such criterion is categorized as the
TABLE 1. Customer churn decision by period and method.

| Setting          | Churn observation criteria | Publishing Information |
|------------------|----------------------------|------------------------|
| Contractual      | Monthly                    | Chen, Yian, et al. (2018) [15] |
| Contractual      | Monthly                    | Mozer, Michael C., et al. (2000) [16] |
| Contractual      | Monthly                    | Daliya, Kiran, and Surbhi Bhatia. (2015) [65] |
| Contractual      | Monthly                    | Bahansen, Alejandro Correa, Djumla Aouada, and Bjorn Ottersten. (2015) [66] |
| Contractual      | Monthly                    | Coussen, Kristof, and Dirk Van den Poel. (2008) [67] |
| Contractual      | Monthly                    | Radosavljevic, Dejan, Peter van der Putten, and Kim Kyllesboch Larsen. (2010) [68] |
| Contractual      | Monthly                    | Gladys, Nicolas, Bart Baesens, and Christophe Croux. (2009) [58] |
| Contractual      | Monthly                    | Hung et al. (2006) [32] |
| Contractual      | Monthly                    | Bezua and van den Poel (2007) [131] |
| Contractual      | Monthly                    | Burez and van den Poel (2008) [132] |
| Contractual      | Monthly                    | Burez, Jonathan, and Dirk Van den Poel (2009) [95] |
| Contractual      | Monthly                    | Madden, Savage, and Coble-Neal (1999) [133] |
| Contractual      | Monthly                    | Gerpoit et al. (2001) [134] |
| Contractual      | Monthly                    | Sco et al. (2008) [135] |
| Contractual      | Monthly                    | Pendhakekar (2009) [136] |
| Contractual      | Monthly                    | Chiu, Tsai, and Ho (2007) [137] |
| Contractual      | Monthly                    | Kim and Yoon (2006) [139] |
| Contractual      | Monthly                    | Ahn et al. (2006) [33] |
| Contractual      | Monthly                    | Wei, Chi-Ping, and I-Tang Chu. (2002) [41] |
| Contractual      | Monthly                    | Neslin, Scott A., et al. (2006) [54] |
| Contractual      | Binary                     | Zhang, Rong, et al. (2017) [11] |
| Contractual      | Binary                     | Dechant, Andrea, Martin Spann, and Jan U. Becker. (2019) [46] |
| Contractual      | Binary                     | Xie et al. (2009) [91] |
| Contractual      | Binary                     | Athanasopoulos (2000) [138] |
| Non-contractual  | Monthly                    | Larivière, Bart, and Dirk Van den Poel. (2004) [36] |
| Non-contractual  | Daily                      | Lee, Eunjoo, et al. (2018) [10] |
| Non-contractual  | Daily                      | Lee, Eunjoo, et al. (2018) [17] |
| Non-contractual  | Daily                      | Tamaddoni Jahromi, Ali, et al. (2010) [20] |
| Non-contractual  | Daily                      | Hadiji, Fahmi, et al. (2014) [42] |
| Non-contractual  | Daily                      | Yang, Wanshan, et al. (2019) [43] |
| Non-contractual  | Binary                     | Buckinx, Wouter, and Dirk Van den Poel. (2005) [96] |

Customer churn can be defined differently for various service characteristics. Generally, customer churn can be decided by customer’s contract revocation, non-use time window period, and service deletion.

churn customer. To conduct this, the customer’s behavioral changed date is counted [96]. When this inactivity or behavioral changed period exceeds the threshold, the customer is regarded as a churn customer. During this process, the period that is set as the threshold of the inactivity date is called the time window [17]. The defining of non-contractual churn has made it possible to infer the probability of the customers who are likely to churn within the certain period. The time window method is frequently used when analyzing activity logs these in a non-contractual situation. When a customer does not use a service for a certain period of time, this method regarding customer as churned. Internet services do not usually delete accounts. Therefore, the Internet service interprets the log-in as prolonged, that is, the retention of the service, and interprets unconnected access for a certain period of time as churn [17]. Fig. 1 schematically illustrates the non-contractual churn case with time window method. The log of Fig. 1 was recorded for 10 weeks. The time window is set to 4 weeks from Week 4 to Week 7. Six users in Fig. 1 showed their activities in each week, and their activities were logged. In the time window period, users A, B, and C without any activity logs from Week 4 to Week 7 are regarded as churn, and the other users D, E, and F with activity logs are regarded as retention.

Churn analysis is usually performed to improve business outcomes. Therefore, in most churn prediction problems, the churn period is defined as a section that can restore customers’ trust. If the time period during which a customer completely churns is selected as a time window, the period for churn definition exponentially increases and it does not provide any gain in terms of business as changing the will of the customers who want to churn is deemed impossible [17]. The contractual mentioned above are close to customers’ complete churn from a service. Therefore, these days majority of the log-based churn prediction problems use the probabilistic method to determine whether customers are churning or not and to give customers incentive to reuse their service.

The criteria for setting the time window are different for each service feature. Yang, Wanshan, et al. (2019) analyzed log data to define the churn period of mobile games, and the analysis result showed that more than 95% of customers did not return when they were absent for 3 consecutive days. They set 3 days as the time window churn period [43]. Lee, Eunjoo, et al. (2018) defined the period during which 75% of customers were continuously unconnected as a churn section by taking into consideration the characteristics of PC game services [17]. After collecting customers’ unconnected periods, they drew a cumulative data graph. They selected the section where more than 75% of customers churned as the time window. Fig. 2 schematically illustrates the cumulative data of consecutive unconnected days collected by Lee, Eunjoo, et al. (2018). According to Fig. 2, the period during which customers were unconnected more than 75% was 14 weeks. Therefore, the time window churn period is 14 weeks.
As described above, there are two customer churn types, which are contractual churn and non-contractual churn. Additionally, there are three churn observation criteria as follows: monthly, daily, and binary. The monthly and daily churn observations are related to the cycle in which the customer’s status is updated in the database. The binary churn observation is acquired by manipulating this database. In general, binary churn is determined by the existence of contract in the contractual settings. In the non-contractual settings, the company defines the customer inactivity features, and when a customer meets the inactivity or disloyal customer feature, the customer is regarded as binary churn [96], [141]. The reason for having multiple ways of defining customer churn is to periodically monitor the customers’ status changes. And through such observation, the expected net business value can be increased by predicting customer churn rates and providing possible churn customers with incentives to retain them from leaving [16], [23].

III. CHURN ANALYSIS IN VARIOUS BUSINESS FIELD
The majority of the early studies on churn were conducted from a management perspective, especially CRM (Customer Relation Management) [30], [31]. CRM churn covers all churn problems that can occur in the process of customer identification, customer attraction, customer retention, and customer development. Modern churn prediction problems are mainly analyzed using log data. A log is trace data that remains when using Internet services. Therefore, the churn prediction models implemented using log data can be used for Internet services in various industries. There are 12 business fields that performed churn prediction. The cases of churn prediction for each business field are summarized in Appendix A.

The telecommunications industry accounts for the majority of previous studies on churn. Telecommunications services have high customer stickiness despite high customer acquisition costs. Therefore, if customer churn is prevented and appropriate incentives are provided, it is of great help in maintaining sales [16], [32]–[34].

The financial and insurance industries also predict customer churn. Zhang, Rong, et al. (2017) stressed the need to build churn prediction models and prevent churn, referring to high customer acquisition costs and high customer values in the insurance industry [11]. Chiang, Ding-An, et al. (2003) mentioned that customer values were high in the online financial market, and created a churn scenario according to the financial product selection and customers’ financial product selection sequence using the Apriori algorithm [35]. Larivière, Bart, and Dirk Van den Poel. (2004), based on the assumption that the customer group was different according to the financial product attribute, demonstrated that the likelihood of churn differed depending on the tendency of customers who selected financial products by measuring the survival time for each product [36]. Zopounidis, Constantin, Maria Mavri, and George Ioannou. (2008) measured the switching rate of financial products, and the survival period of customers for each product to discover attractive products [37]. Here, as the survival period is short, churn occurs more frequently, which is used as an indicator to measure the need to supplement financial products. Glady, Nicolas, Bart Baesens, and Christophe Croux. (2009) measured the customer lifetime values and the decrease in expected earnings over time as an indicator corresponding to customer loyalty [38]. During this process, machine learning was used to calculate the churn rate which was used to estimate the customer lifetime values.

Later on, studies on churn have been actively conducted in the gaming field as in the telecommunication field. These services have a fast cycle of customer inflow and churn because of mass competition. However, if a single service is run for a long time, the service competition intensifies and the Customer Acquisition Cost (CAC) tends to increase [16], [39], [128]. As the CAC gets larger, the technology to predict and prevent churn becomes more crucial. Viljanen, Markus, et al. (2016) applied the survival analysis to mobile games and calculated the churn rate, similar to the churn prediction of financial services [40]. The game sector actively uses machine learning techniques when conducting research on churn because of the large volume of log data [10], [42], [43]. Milosevic, Milos, Nenad Živic, and Igor Andjelkovic. (2017) created a model predicting churn in the study on game churn, gave churn prevention incentives by finding out and dividing probable churn customers into A/B groups, and demonstrated actual effects statistically [44]. Runge, Julian, et al. (2014) conducted a similar study, and revealed that existing customers with a high possibility of churn had a higher marketing response rate when compared to general marketing targets [45].

Furthermore, the music streaming service field even held a competition to build a prediction model, and research on churn was also conducted in the Internet service and...
newspaper subscription fields. The newspaper subscription and music streaming service offer fixed-rate services, and customer churn is consistent with the contract renewal period. On the other hand, because the Internet service goes into an inactive state as customers wish, contract renewal takes place nearly-real-time. Research on churn prediction was also conducted in online dating, online commerce, Q&A services, and social network-based services [46].

There were some studies which approached customer churn from a psychological perspective. Borbora, Zoheb, et al. (2011) analyzed that customers churned when their motivation to use games changed by combining the motivation theory with customers using MMO RPG games [47]. Yee, Nick. (2016) surveyed approximately 250,000 gamers, and showed that customers’ attitudes toward games were clustered by country, race, and age [48].

In the marketing field, Gladly, Nicolas, Bart Baesens, and Christophe Croux. (2009) used the features from a marketing perspective such as RFM (Recency, Frequency and Monetary) and CLV (Customer Life time Value) for churn prediction [38].

Studies on churn prediction were conducted in the human resources and energy fields although they were minority. Saradhi, V. Vijaya, and Girish Keshav Palshikar. (2011) conducted research on churn to reduce retraining costs when employees churned and to prove employee value in the human resources field [50]. Moeyersoms, Julie, and David Martens. (2015) estimated whether customers would churn to another energy supplier based on energy data and socio-demographic data provided to customers [51].

IV. CUSTOMER CHURN LOSS
Customer churn behavior is quantitative. However, it is difficult to directly relate customer churn to a decrease in sales. Therefore, of the studies on churn prediction, there is a study that introduces a method of calculating the loss of a single customer. In this way, we can calculate the value of a churn prevention model by multiplying the loss cost of one customer by the number of people who are prevented from churn with a churn prediction algorithm.

A. CUSTOMER ACQUISITION COST (CAC)
The customer acquisition cost (CAC) is the total cost that is spent until a customer is convinced of a service. CAC can be calculated by simply dividing all cost spent on acquiring customers, marketing campaign for example, by the number of customers acquired in the period the money was spent. The company measures the cost of acquiring a customer with the CAC, and that CAC cost is the minimum value to operate service makes from a customer with Return on Investing (ROI). The CAC occurs mainly through marketing activities. If a customer churns from a service, the company will have to recruit another customer by spending the CAC to maintain the service. According to the study conducted by Mozer, Michael C., et al. (2000), the retained customers are known to provide a better return on investment than the newly recruited customers through the CAC [16]. In such way, if a customer who is likely to churn the service in the near future can be inferred through the churn prediction, the basis for measuring a suitable incentive for the customer while minimizing the CAC can be established. In addition, by multiplying the number of customers planning to churn by the incentive cost to be provided to those customers, the cost of business loss that can incur when the customers are not retained can be calculated through the churn prediction model [45], [53]. Therefore, some studies measured the CAC, and calculated it as the loss incurred when a customer churned [8], [31], [72].

B. CUSTOMER LIFETIME VALUE (CLV)
The customer lifetime value (CLV) is the cost that a customer expects to pay when acquiring a customer. This reason why this cost is important is that the CLV is the expected earnings from the customer’s use of a service when acquiring a new customer, and the CLV is a useful indicator for setting the upper bound when calculating customer-related costs. Marketing costs and incentives provided for customers who are going to churn are typical examples of customer-related costs [52]. The efficiency of a retained customer value is usually calculated with the CLV. A retained customer represents a customer who the churn prediction model predicted to churn in the near future but survived after receiving incentives. This is because the cost can vary depending on the company’s policy and marketing timing for the CAC method. There are multiple studies on the method of calculating the CLV and that apply the CLV to churn models. Verbraken, Thomas, Wouter Verbeke, and Bart Baesens. (2012) and Neslin, Scott A., et al. (2006) proposed formulas for calculating the net profit using churn rate, CAC, CLV, fixed operating cost, and incentive cost [54], [55]. Additionally, the same approach has been taken by Fader, Peter S., and Bruce GS Hardie. (2009) as well [149]. In the formula for deriving CLV, the survival rate (retention) γ for the customer’s time period t should be derived first. γ can be derived through the probabilistic distribution as well. When γ denotes a retention rate, 1 − γ expresses a churn rate. In this case, the expected survival time of a customer can be expressed as \frac{1}{1−γ}. Lastly, assuming that the profit contribution cost per customer (customer value) for t period is expressed as m, the CLV can be obtained by mp, where m may be different for each customer segment. In the case of calculating m value of new acquisition customer, m is derived by dividing the net profit from active customers for time t by the number of active customers. Further, the contribution of a specific segment can be calculated by dividing the net profit from activated specific segment customers for t period by the number of activated segment customers during t period. In the case where the customer value is discounted during the time t, the discount rate is defined as \delta, and the discount value during the time t is expressed as \frac{mp}{(1+\delta)^t}. Ultimately, the CLV having discount term for a given time t can be expressed as \frac{mp}{(1+\delta)^t}. Since the CLV should include the concept of lifetime, time t can be
generalized as follows.

\[
CLV = \sum_{t=0}^{T} m \frac{y^t}{(1 + d)^t}
\]

As a way of calculating the loss of employee churn, Saradhi, V. Vijaya, and Girish Keshav Palshikar. (2011) calculated churn rates, and by using the method of multiplying customer value by the remaining survival time, the authors calculated the projected value in future time of churned employees who failed to fulfill the CLV [50]. Based on the formula for deriving net cashflow using survival time parameters suggested by Reinartz, Werner J., and Vijay Kumar. (2000), the study conducted by Glady, Nicolas, Bart Baesens, and Christophe Croux. (2009) used the approach of multiplying individual cashflows for entire product to calculate the CLV in the retail service field [38], [146].

V. FEATURE ENGINEERING

Churn is generally related to customers’ last time activity. However, predicting churn and compensating for it with the last log of service usage do not change the overall service usage patterns of customers. Therefore, some studies show that short-term prediction and monetary rewards soon leads to another churn [10], [45]. Therefore, studies have emerged in recent years to develop other features that are as important as the last log the customer left before churning, or to discover potential churners by reprocessing the time series features.

A. DEVELOPING NEW FEATURE

Sifa, Rafet, Christian Bauckhage, and Anders Drachen. (2014) diagrammed the related signs leading to churn, and grouped the features corresponding to each diagram and managed them in Game field [56]. This study focused on detecting signs that led to churn rather than building churn models and comparing performance. They linked unmeasurable numbers such as key indicators of services, the number of complaints and psychological fluctuations to service features so as to measure them, and utilized them for the research on churn prediction.

Yang, Wanshan, et al. (2019) judged that the probability of churn would increase when the regularity of customers’ behavior using services was broken [43]. They added the change in the customer service playtime distribution as a feature, and maintained that the feature was of a great help in estimating churn by creating a machine learning model. Hadiji, Fabian, et al. (2014) and Yang, Wanshan, et al. (2019) predicted churn using KPI features [42], [43]. They used the indicators used in business administration as churn prediction features since churn was related to management indicators.

Runge, Julian, et al. (2014) associates the value of service goods possessed by customers with customer churn. This study intensively used features related to assets (e.g., reserves, items) in customers’ services [45]. Paid goods, free goods, last purchase, and last purchase date, and so on were used as assets. They assumed that as the user had more goods to use in the service, the opportunity cost became larger, which would be a major indicator of churn.

In the finance field, Chu, Tsai, and Ho (2007) used customer demographic CRM features and business branch relationships to conduct churn prediction [137]. They have predicted customer churn using the customer information such as gender, zip code, and customer’s industry code and the service provider information such as tenure, time of service suspended, and average invoice. In order to predict customer churn of Pay-TV services, Burez, Jonathan, and Dick van den Poel (2008) selected customer behavioral loyalty features and combined them with CRM features [132]. In addition to using service information such as payment type and contract expiration month, the authors used demographic CRM features including customer’s age, province, and customer type and additional features for classifying disloyal customers—namely bad payment behaviors, number of notices to pay, and number of deactivation of the device—in order to conduct churn prediction.

Logs remain in the majority of services. However, not all logs are helpful in estimating churn. Mozer, Michael C., et al. (2000) maintained that in general the indicators that measured the quality of services were good data for estimating churn in telecommunication field [16]. Dror, Gideon, et al. (2012), by collecting responses such as like and dislike from the Internet service they ran, used them to predict churn. They explained that customer satisfaction was expressed as an emotional expression, which was a direct expression of service satisfaction [57].

Fader, Peter S., Bruce G S Hardie, and Ka Lok Lee. (2005) and Glady, Nicolas, Bart Baesens, and Christophe Croux. (2009) applied the RFM method, which is used for selecting loyal customers in the marketing field, to the churn prediction [38], [49]. The RFM stands for recency (latest service transaction), frequency (service transaction frequency) and, monetary (customer’s purchase size). These features are used to extract customers who carry business values. In terms of marketing, the RFM features are generally used to classify customers into five groups based on the RFM scores [140]. However, in the above two studies, the authors characterized RFM features as the important features that can derive the net business values of customer-service relationship and conducted churn prediction based on the premise that loyal customers have higher service stickiness and switching cost. On the other hand, to conduct a customer churn prediction, Tamaddoni Jahromi, Ali, et al. (2010) applied RFM features to the telecommunication field and added about 12 new features, including the latest telecommunication service subscription period, call frequency, total call cost, etc. [20]. The authors also used mobile carrier-specific features including call time, number of incoming or outgoing calls, and total talk time between specific customers to conduct churn prediction. Further, Wei, Chih-Ping, and I-Tang Chiu. (2002) have utilized RFM features in the telecommunication field as well [41]. By using the time length between the contract starting date and termination date, the authors set frequency...
of service use as the recency feature, payment type as the frequency feature, and payment type as the monetary feature. Additionally, as for the mobile carrier-specific feature, the authors derived an influence feature, which indicates the number of distinct receivers the customer called in the outgoing call list. Buckinx, Wouter, and Dirk Van den Poel. (2005) applied RFM features to the retail field to conduct churn prediction [96]. The authors used features of customer’s recent purchase or consumption time of the day (Recency), number of purchases (Frequency), and the amount of spending (Monetary). As for the retail service-specific features, metadata such as customer-supplier relationship, buying categories, mode of payment, brand purchase behavior, and usage of promotions were used.

B. FEATURE MODIFICATION

Pure log data includes all truthful customer information. However, churn prediction tends to be more accurate when processing raw logs. This is because service indicators are generally sparse and have many outliers, and the data distribution is skewed to one side. Therefore, it is necessary to extract important information using an appropriate feature engineering technique for churn prediction.

In general, there have been few papers mentioning feature engineering know-how about churn. However, Zhang, Rong, et al. (2017) shared useful information when building an algorithm to predict churn from log data [11].

1) ONE-HOT ENCODING

One-hot encoding is primarily a feature engineering method for nominal categorical data. In order to implement machine learning on categorical data without using a tree-based algorithm, the data must be converted to numerical data. The algorithm used in this instance is called the one-hot encoding. One-hot encoding is expressed by a group of bits among which the legal combinations of values are only those with a single high (1) bit and all the others low (0) bits. One-hot encoding on categorical data in this way produces an orthogonal feature space for each category. Although alternative methods, such as numerical encoding or binary encoding, exist for processing categorical data, each method has its drawbacks: Numerical encoding breaks the nominal nature inherent in categorical data and results in encoding with linearity between categories, likewise binary encoding produces distance between categories. These changes have the side effect of the model learning unintentional continuity. Therefore, categorical data should be modified into one-hot encoding features.

2) BUCKETING

Outliers are problematic for service data. Bucketing (also called Discrete binning or Data binning) can be used for both categorical and continuous features. For features that are too sparse to be used in the model due to the large variance because the feature value category is too wide, bucketing is a technique that makes these features into categorical features. For example, a positive integer feature \( x \) can be modified into three discrete bucket: 1 (if \( 0 \leq x < 5 \)), 2 (5 \( \leq x < 10 \)), and 3 (if 10 \( \leq x \)).

3) DATA IMPUTATION

Data imputation is the process of replacing missing data with substituted values. It is recommended to fill the missing data by any means. Nimmagadda, Sravya, Akshay Subramaniam, and Man Long Wong. (2017) argued that filling the missing data improved performance instead of dropping them [58]. Sifa, Rafet, et al. (2015) improved performance by using a semi-supervised learning technique since there was no sufficient data to solve prediction problems [52]. Data imputation techniques may help to solve data imbalances and make the best of the information used in the model.

4) NORMALIZATION

Normalization is data pre-processing technique for stability in several training machine learning algorithms. This process scale individual samples to have unit norm. The distribution of service data is generally skewed to one side. Without normalization, the machine learning model generally have to select a very small learning rate when searching for the optimum, resulting in a long training time. Normalized features can achieve rapid model training while using relatively large learning rates compared to the initial data. Nimmagadda, Sravya, Akshay Subramaniam, and Man Long Wong. (2017) used techniques such as log-normalization or quantile normalization to build a prediction model other than 0-1 normalization [58].

5) FEATURE EMBEDDING

Schweidel, David A., and George Knox. (2013) proposed a parsimonious model by integrating customer behavior data into latent attrition models in order to provide direct marketing target selection [142]. This model enables dense embedding of sparse customer data by extracting latent attributes. In general, when the customer behavior is stored as log data in gaming, Internet service, and telecommunication service fields, the customer data is often found in the form of high-dimensional sparse data. Hence, in the past, the customer behavior data was simplified through a bucketing process. However, with the emergence of deep learning techniques of machine learning, it has become possible to manage time-dependent high-dimensional sparse data. Unlike the explicit methods such as bucketing, a deep learning algorithm can learn that customers’ latent behavior who are about to churn in a end-to-end way. Moreover, a deep learning algorithm enables generating latent features by compressing long-term features that is a new technique of feature embedding. An autoencoder is one of the example techniques to this. Autoencoder is trained based on the encoding and decoding process, where latent vectors are generated during the process. The latent features generated during this process compresses high-dimensional sparse data into low-dimensional dense data. It has been suggested that the model
that uses these obtained vectors as input of the fully connected networks provides better prediction performance than the traditional model that uses the sparse features as is. For example, to predict the future demand of Uber customers, Zhu, Lingxue, and Nikolay Laptev. (2017) improved the performance of the prediction model by compressing sparse data through a long-short-term memory autoencoder and then concatenating the data with fully connected neural networks (FCNNs) [143]. This trend is found in the churn prediction as well. Lee, Eunjoo, et al. (2018) claimed that in the customer churn prediction competition hosted by the authors, the winning team showed significantly better performance than other teams by utilizing the autoencoder to compress the sparse data [10]. Zhang, Rong, et al. (2017) also used latent feature modification for churn prediction. They classified features into a memorization feature and a generalization feature depending on data attributes [11]. The memorization feature refers to data that is likely to be a latent feature among time series data. These types of data slowly reveal the characteristics of the churn/no-churn group usually over a period of time. This data is modified by a deep learning model LSTM, embedding, or autoencoder, and then processed by converting them into a dense-low dimension latent feature. As opposed to the memorization feature, the generalization feature exhibits the characteristics of the churn/no-churn group with only short-term attributes. This feature can represent the attributes of the churn/no-churn group with a short-term section or value itself. The generalization feature data predicts results with shallow machine learning models such as logistic linear regression. The churn feature that combines the memorization feature and the generalization feature predicts churn by combining deep learning models and traditional machine learning techniques.

C. DEALING WITH IMBALANCED DATA

In a stable service, the number of churned customers generally account for a small proportion compared to the number of retained customers. For example, suppose that 96 percent of the data is composed of retained customers and 4 percent is churned customers. If a prediction model is trained using this data and it always outputs only the results that indicate customers being retained, the model would maintain a 96 percent accuracy. In this case, although the accuracy of the model may be said to be high, the model would not be effective in identifying characteristics of churned customers. Thus, the most ideal method is adjusting the churned data and retained data to have a similar proportion. This is because a balanced dataset composed of churned and retained customers has higher noise tolerance than an imbalanced dataset, hence it is more likely to be able to obtain decision boundaries for minor groups, which in this case denotes the churned customer dataset [91]. Meanwhile, there could be an issue where the prediction model may not have enough data for training if a simple undersampling method is used to obtain a balanced dataset. To address this, Burez, Jonathan, and Dirk Van den Poel. (2009) utilized a method called the CUBE method to improve the churn prediction performance. [95]. In addition, Amin, Adnan, et al. (2016) obtained the IBM telecommunication dataset and applied the oversampling method on the imbalanced dataset. Subsequently, the authors used the generated balanced dataset to train a churn prediction model [144]. The authors also compared and analyzed the performance of the churn prediction models that implemented widely used oversampling methods. According to the study, the Mega-Trend Diffusion Function (MTDF) method provided the highest accuracy when compared with other techniques such as Synthetic Minority Oversampling Technique (SMOTE), Adaptive Synthetic Sampling Approach (ADASYN), Majority Weighted Minority Oversampling Technique (MWMOTE), Immune Centroids Oversampling Technique (ICOTE) and couples Top-N Reverse K-Nearest Neighbor (TRkNN) algorithm. Gui, Chun. (2017) applied undersampling, oversampling, and SMOTE methods to an imbalanced dataset of telecommunication field and compared the performance of the derived churn prediction models [145]. In the study, the author suggested that the SMOTE sampling technique provided the best prediction performance. The use of undersampling technique when dealing with churn dataset has disclosed the possibility of discarding useful information that expresses the characteristics of retained customers. On the contrary, the use of oversampling technique has disclosed the possibility of raising the overfitting issue due to replicating insufficient variance size of the oversampled churned group data.

VI. CHURN PREDICTION MODELS

A. BUILDING CHURN PREDICTION MODELS

There are four domains as a method of building a churn prediction model. They are traditional machine learning, statistics, graph theory, and deep learning. In Appendix B-A, we summarized the papers that built churn prediction models based on these four criteria.

The boundary is blurring between the above four disciplines in recent times. However, Breiman, Leo. (2001) divided statistics and machine learning into a stochastic data model and an algorithmic model, which influenced the way that Matthew Stewart. (2019) and Bzdok, D., Altman, N. and Krzywinski, M. (2018) divided statistics and machine learning [60]–[62]. Witten, Ian H., and Eibe Frank. (2002) described that machine learning techniques had developed with data mining since the advent of computers while statistics had focused on mathematics-based hypothesis tests [63].

In statistics, probability models have been mainly used for conducting churn predictions. In particular, probability models have been traditionally used for customer-base analysis [149]. In a customer-base analysis, churn rate is applied to the survival time estimation when calculating the CLV. Figure 3 illustrates that the appearance of churn prediction algorithms by year. CLV prediction algorithm is combined with calculating customer expected revenue and churn rate prediction model. To calculate CLV within the contractual settings,
a shifted-Beta-Geometric (sBG) model is used. The sBG model uses beta distribution to make shifts for every instances of change in time \( t \) in order to fit the retention rate. Accordingly, the sBG model allows continuous interpretation of the conditions in which the customer retention is determined to be discrete-time contractual in a contractual service [150].

In the non-contractual settings, the repeat-buying behavior of customers have been previously expressed through negative binomial distributions (NBD) [151]. Further, the distribution of churn used gamma mixture of exponential, which is also known as the Pareto (of the second kind) distribution. By combining the buyer behavior and survival distribution, the CLV can be calculated. and this method is simply referred to as the Pareto/NBD model. The Pareto/NBD method has been actively used as a probability model for deriving the CLV until recently [19], [147], [148]. As another method, the beta-geometric/beta-binomial (BG/BB) model can be used. In this model, the beta-geometric model fits the retention rate and the beta-binomial distribution fits the consumer purchasing behavior [149]. In the non-contractual settings, customer churn tends to have a characteristic of continuous probabilistic. In the non-contractual settings, customer churn is not easy to define and trace. To define customer churn in the non-contractual settings, researchers used in time-series modification techniques such as grouping customer id to make tidy data or calculate behavioral variances [114]. Utilizing the features produced from this processed data, researchers begin predicting customer churn. The statistical model used here is based on statistical inference and hypothesis testing, and survival analysis with hazard methods are used to build churn prediction models. Machine learning techniques also began to be utilized in customer churn for non-contractual settings. Compared to statistical methods, machine learning techniques have robust non-linear relationships between features and can learn heterogeneous effect when given diverse features. The graph theory identifies churn as a mathematical relationship. It configures graph attributes by feature and by customer, and expresses their relationship as edges. Once a graph is built, it searches churning customers through the graph correlation analysis. Recently, deep learning techniques have emerged as a method of predicting customer churn. Deep learning is an intense extension of machine learning with neural network algorithms. However, deep learning also has many variances and tend to be classified separately from conventional machine learning. Deep learning techniques often used for predicting customer churn mainly involves training via sparse customer data that have been densified, or a fully-connected neural network produced from extracting the latent vector from the autoencoder and concatenating those features with static features.

The deep learning model is a relatively recent analysis method of predicting churn. According to Goodfellow, Ian, Yoshua Bengio, and Aaron Courville. (2016), deep learning is part of machine learning [64]. However, because its academic significance has recently grown, it has established itself as a single academic field. This is true for building churn prediction analysis models. Lee, Eunjo, et al. (2018) disclosed that a model using deep learning predicted customer churn with a higher probability than a traditional machine learning model for game churn prediction analysis [10]. Fig. 4 schematically represents the model of the team that won this competition. They summarized the features using the memorization and generalization techniques described in feature modification, and increased churn predictability by combining the deep learning model and the traditional machine learning model. Zhang, Rong, et al. (2017) compared the traditional machine learning model and the deep learning model in customer...
churn prediction problems in the insurance industry [11]. Fig. 5 is a schematic diagram of the Deep and Shallow model they built. Zhang, Rong, et al. (2017) classified features to be applied to the deep learning model and processed them, and then combined the results. In the study, they compared the deep learning-based churn prediction method they developed with the traditional machine learning-based churn prediction algorithm. The Deep and Shallow model they built showed excellent churn prediction performance compared to other models. In this study, although deep learning is part of machine learning, it is used as a new breakthrough algorithm for churn prediction problems. Appendix B-A shows the classification of studies on churn based on churn prediction algorithms.

B. CHURN PREDICTION MODELS IN VARIOUS BUSINESS FIELD

Table 2 shows a summary of techniques that are classified by business field. We were able to confirm that preferred modelling techniques were different depending on the business field. Businesses with dense log data and easy access to customer information, such as the games and telecommunications industries, are applying relatively many deep learning techniques using big data, which is a fast trend. As for the financial and insurance sectors, since the log data is relatively small and the information obtained from customers does not change to a great degree, there are many statistical approaches using traditional machine learning models or survival analysis. The reason why the preferred model for each business field is different is that the types and cycles of log data used for each business are different. Apart from this, it seems to be different since the churn model that best interprets the relevant log data is applied.

C. PERFORMANCE EVALUATION

In general, the performance evaluation algorithms of machine learning model developed for churn prediction use the area under the curve (AUC) of receiver operating characteristic (ROC) curve or the lift. The ROC curve is drawn by plotting sensitivity values on the y-axis and false positive rate on the x-axis. A ROC curve is a very robust measurement criterion that measures classifiers independent of class distribution and misclassification error cost. In this way, x-axis denotes the proportion of non-churn cases that were incorrectly classified as churn, and y-axis is defined as the portion of churn cases that were classified correctly [121], [136]. Thus, the AUC close to 1 indicates that the churn prediction model accurately distinguishes difference between the characteristics of churn customers and non-churn customers [13], [45], [55], [95]. On the other hand, some churn studies have often used a top 10% decline lift performance metric [13], [67]. Lift is a performance measure obtained by dividing baseline lift by the response for each fraction. When using top 10% decline lift as in the above reference studies, the customer list sorted in the descending order based on the prediction rate is divided into ten fractions. Subsequently, the lift values for each fraction are derived and the descending speed of the curve is observed. Additionally, top-decile lift technique is also often used as it enables allocating marketing budgets proportionally to customers who are more likely to churn as predicted by the churn prediction model [95].

VII. CONCLUSION

In this study, we compared the churn prediction analysis techniques using log data. Churn analysis is used in the fields of Internet services and games, insurance, and management. Research on churn prediction usually begins to improve business outcomes. Therefore, the time window is used to select potential churning customers rather than measuring a customer’s complete churn. Loss costs for customer churn are calculated by CAC or CLV. In the past, when predicting customer churn, researchers used survival analysis or time series analysis using statistics, graph theory, and traditional machine learning algorithms. Churn prediction analysis using deep learning algorithms has recently emerged. Deep learning algorithms have been found to outperform other algorithms. This is likely due to large quantities of customer log data being collected via computers and the churn prediction model utilizing the entire set of this acquired data to make
| Industry       | Application                                          | Publishing Information                          | Industry       | Application                                          | Publishing Information                          |
|---------------|-----------------------------------------------------|-------------------------------------------------|---------------|-----------------------------------------------------|-------------------------------------------------|
| Game          | Churn Prevention                                    | Runge, Julian, et al. (2014) [45]               | Game          | Game Design                                         | Drachen, Anders, Magy Seif El-Nasr, and Alessandro Canossa, eds. (2013) [72] |
| Game          | Churn Prevention                                    | Periáñez, África, et al. (2016) [14]            | Game          | Churn Prevention                                    | Kim, Seungwook, et al. (2017) [69]              |
| Game          | Churn Prevention, Customer Clustering               | Bindewald, Jason M., Gilbert L. Peterson, and Michael E. Miller. (2016) [70] | Game          | Churn Prevention                                    | Ben Lewis-Evans. (2012) [71]                   |
| Game          | Churn Prevention                                    | Lee, Eunjo, et al. (2018) [10]                  | Game          | Churn Prevention                                    | Bertens, Paul, Anna Guitart, and África Periáñez. (2017) [73] |
| Game          | Churn Prevention                                    | Tamassia, Marco, et al. (2016) [75]             | Game          | Churn Prevention                                    | Hadiji, Fabian, et al. (2014) [42]             |
| Game          | Churn Prevention                                    | Bastiaan van der Palen. (2017) [76]             | Game          | Churn Prevention, Customer Profitability            | Sifa, Rafet, et al. (2015) [52]                |
| Game          | Churn Prevention, Customer Profitability            | Lee, Eunjo, et al. (2018) [17]                  | Game          | Game Design                                         | Fields, Tim, and Brandon Cotton. (2011) [77]    |
| Game          | Churn Prevention                                    | Sifa, Rafet, Christian Bauckhage, and Anders Drachen. (2014) [56] | Game          | Churn Prevention                                    | Yang, Wanshan, et al. (2019) [43]             |
| Game          | Game Design                                         | Fields, Tim. (2014) [3]                         | Game          | Churn, Customer Clustering                          | Kawale, Jaya, Aditya Pal, and Jaideep Srivastava. (2009) [78] |
| Game          | Churn Prevention                                    | Jeff Grubb. (2014) [39]                         | Game          | Churn Prevention                                    | Kristensen, Jeppe Theiss, and Paolo Burelli. (2019) [79] |
| Game          | Churn Prevention                                    | Nozhnin, Dmitry. (2012) [74]                   | Game          | Churn Prevention                                    | Guitart, Anna, Pei Pei Chen, and África Periáñez. (2019) [80] |
| Game          | Churn Prevention                                    | Viljanen, Markus, et al. (2017) [81]            | Game          | Churn Prevention                                    | Lee, eunjo. (2019) [82]                       |
| Game          | Churn Prevention                                    | Bauckhage, Christian, et al. (2012) [83]       | Game          | Churn Prevention                                    | Shores, Kenneth B., et al. (2014) [84]         |
| Game          | Churn Prevention                                    | Debeauvais, Thomas, et al. (2011) [85]          | Game          | Churn, Customer Profitability                       | Milošević, Miloš, Nenad Živić, and Igor Andjelković. (2017) [44] |
| Game          | Churn Prevention                                    | Viljanen, Markus, et al. (2016) [40]            | Game, Marketing | Churn Prevention                                    | Castro, Emiliano G., and Marcos SG Tsuzuki. (2015) [86] |

Churn analysis in various business field.

Some of the papers introduced in VI. Churn Prediction Models of this paper used deep learning for churn prediction with data timestamps in the order of seconds or with vast amounts of customer log data in total. In such case, feature engineering techniques for processing logs have a significant effect on model performance enhancement. Unlike other modeling techniques, the deep learning model is capable of converting high dimensional sparse log data into low dimensional dense features by embedding time series features. Also the deep learning model can learn cus-
### TABLE 3. (Continued.) Churn analysis in various business field.

| Domain          | Churn Prevention, Customer Clustering | Churn Prevention, Customer Profitability | Insurance | Marketing | Management, Customer Profitability | Management, Customer Behavior Analysis | Retails |
|------------------|---------------------------------------|-----------------------------------------|-----------|----------|------------------------------------|--------------------------------------|---------|
| Game, Psychology | Borbora, Zoheb, et al. (2011) [47]     | Game, Psychology                        | Yee, Nick. (2016) [48]               |          |                                    |                                      |         |
| Game, Patent     | Wolters, Hans, Jim Baer, and Girish Keswani. (2014) [87] | Game, Gamble                           | Coussement, Kristof, and Koen W. De Bock. (2013) [88] |          |                                    |                                      |         |
| Finance          | Glad, Nicolas, Bart Baesens, and Christophe Croux. (2009) [38] | Finance                                | Zopounidis, Constantin, Maria Mavri, and George Ioannou. (2008) [37] |          |                                    |                                      |         |
| Finance          | Chiang, Ding-An, et al. (2003) [35]    | Finance                                | Lariviére, Bart, and Dirk Van den Poel. (2004) [36] |          |                                    |                                      |         |
| Finance          | Xie, Yaya, et al. (2009) [91]          | Finance                                | Anil Kumar, Dudyala, and Vadlamani Ravi. (2008) [92] |          |                                    |                                      |         |
| Finance          | Nie, Guangli, et al. (2011) [93]       | Finance                                | Ali, Ozden Gür, and Umut Arittürk. (2014) [94] |          |                                    |                                      |         |
| Finance          | Evermann, Joerg, Jana-Rebecca Rehse, and Peter Fetteke. (2017) [18] | Finance                                | Athanassopoulos (2000) [138]          |          |                                    |                                      |         |
| Finance          | Lim, Tong-Ming, and Angela Siew Hoong Lee. (2017) [152] | Finance, Telecom., Newspapers          | Burez, Jonathan, and Dirk Van den Poel. (2009) [95] |          |                                    |                                      |         |
| Finance          | De Bock, Koen W., and Dirk Van den Poel. (2011) [103] | Finance, Telecom., Commerce            | Verbraken, Thomas, Wouter Verbeke, and Bart Baesens. (2012) [55] |          |                                    |                                      |         |
| Finance          | Chu, Tsai, and Ho (2007) [137]         | Insurance                              | Morik, Katharina, and Hanna Köpcke. (2004) [98] |          |                                    |                                      |         |
| Insurance        | Hur, Yeon, and Sehun Lim. (2005) [99]  | Insurance                              | Risselada, Hans, Peter C. Verhoef, and Tammo HA Bijnfelt. (2010) [100] |          |                                    |                                      |         |
| Insurance        | Zhang, Rong, et al. (2017) [11]        |                                        |           |          |                                    |                                      |         |
| Marketing        | Xie, Hanting, et al. (2015) [127]      | Marketing                              | Fader, Peter S., Bruce GS Hardie, and Ka Lok Lee. (2005) [49] |          |                                    |                                      |         |
| Management       | Ngai, Eric WT, Li Xiu, and Dorothy CK Chau. (2009) [30] | Management                            | Lejeune, Miguel APM. (2001) [31]       |          |                                    |                                      |         |
| Management       | Butgereit, Laurie. (2020) [89]         | Management, Patent                     | Wright, Christine. (2003) [90]         |          |                                    |                                      |         |
| Management       |                                        |                                        |                                        |          |                                    |                                      |         |
| Retails          | Clemente, M., V. Giner-Bosch, and S. San Matías. (2010) [97] | Retails                               | Buckinx, Wouter, and Dirk Van den Poel. (2005) [96] |          |                                    |                                      |         |
| Business Field          | Churn Analysis                  | Authors                                      |
|------------------------|---------------------------------|----------------------------------------------|
| Newspapers             | Prevention, Customer Profitability | Neslin, Scott A., et al. (2006) [54]         |
| Newspapers             | Prevention                      | Coussement, Kristof, Dries F. Benoit, and Dirk Van den Poel. (2010) [101] |
| Music Streaming Service| Prevention                      | Nimmagadda, Sravya, Akshay Subramaniam, and Man Long Wong. (2017) [58] |
| Internet Service       | Prevention                      | Dechant, Andrea, Martin Spann, and Jan U. Becker. (2019) [46] |
| Internet Service       | Prevention                      | Internet Service                             |
| Internet Commerce      | Prevention                      | Yu, Xiaobing, et al. (2011) [106]            |
| Energy                 | Prevention                      | Moeyersoms, Julie, and David Martens. (2015) [51] |
| Pay TV                 | Prevention                      | Pay TV                                       |
| Telecom.               | Prevention                      | Dalvi, Preeti K., et al. (2016) [107]         |
| Telecom.               | Prevention                      | Ge, Yizhe, et al. (2017) [108]               |
| Telecom.               | Prevention, Customer Profitability | Mozer, Michael C., et al. (2000) [16]       |
| Telecom.               | Prevention, Customer Profitability | Daihya, Kiran, and Surbhi Bhatia. (2015) [65] |
| Telecom.               | Prevention, Customer Profitability | Verbraken, Thomas, Wouter Verbeke, and Bart Baesens. (2014) [4] |
| Telecom.               | Prevention                      | Dasgupta, Koustrup, et al. (2008) [110]      |
| Telecom.               | Prevention                      | Tsai, Chih-Fong, and Yu-Hsin Lu. (2009) [111] |
| Telecom.               | Prevention                      | Hudaib, Amjad, et al. (2015) [112]           |
|                         |                                 | Coussement, Kristof, and Dirk Van den Poel. (2008) [67] |
|                         |                                 | Chen, Yian, et al. (2018) [15]               |
|                         |                                 | Ngonmang, Blaise, Emmanuel Viennet, and Maurice Tchuente. (2012) [105] |
|                         |                                 | Madden, Savage, and Coble-Neal (1999) [133]  |
|                         |                                 | Tamaddoni, Ali, Stanislav Stakhovych, and Michael Ewing. (2016) [102] |
|                         |                                 | Burez and van den Poel (2007) [131]         |
|                         |                                 | Saradhi, V. Vijaya, and Girish Keshav Palshikar. (2011) [50] |
|                         |                                 | Hung, Shin-Yuan, David C. Yen, and Hsiu-Yu Wang. (2006) [32] |
|                         |                                 | Ahn, Jae-Hyeon, Sang-Pil Han, and Yung-Seop Lee. (2006) [33] |
|                         |                                 | Wei, Chih-Ping, and I-Tang Chiu. (2002) [41] |
|                         |                                 | Vafiadis, Thanasis, et al. (2015) [8]       |
|                         |                                 | Baumann, Annika, et al. (2015) [109]        |
|                         |                                 | Au, Wai-Ho, Keith CC Chan, and Xin Yao. (2003) [34] |
|                         |                                 | Tamaddoni Jahromi, Ali, et al. (2010) [20]  |
|                         |                                 | Öskarsdóttir, María, et al. (2018) [113]    |
 TABLE 3. (Continued.) Churn analysis in various business field.

| Telecom. | Churn Prevention | Qian, Zhiguang, Wei Jiang, and Kwok-Leung Tsui. (2006) [114] | Telecom. | Churn Prevention | Wangperawong, Artit, et al. (2016) [115] |
|----------|------------------|------------------------------------------------------------|----------|------------------|------------------------------------------|
| Telecom. | Churn Prevention | Radosavljevik, Dejan, Peter van der Putten, and Kim Kyllesbech Larsen. (2010) [68] | Telecom. | Churn Prevention | Richter, Yossi, Elad Yom-Tov, and Noam Sionim. (2010) [53] |
| Telecom. | Churn Prevention | Hadden, John, et al. (2006) [117] | Telecom. | Churn Prevention | Nath, Shyam V., and Ravi S. Behara. (2003) [118] |
| Telecom. | Churn Prevention | Lemmens, Aurélie, and Christophe Croux. (2006) [119] | Telecom. | Churn Prevention | Huang, Ying, and Tahar Kechadi. (2013) [120] |
| Telecom. | Churn Prevention | Huang, Bingquan, Mohand Tahar Kechadi, and Brian Buckley. (2012) [121] | Telecom. | Churn Prevention | Idris, Adnan, Muhammad Rizwan, and Asifullah Khan. (2012) [122] |
| Telecom. | Churn Prevention | Dierkes, Torsten, Martin Bichler, and Ramayya Krishnan. (2011) [123] | Telecom. | Churn Prevention | Kim, Kyungok, Chi-Hyuk Jun, and Jaewook Lee. (2014) [124] |
| Telecom. | Churn Prevention | Keramati, Abbas, et al. (2014) [125] | Telecom. | Churn Prevention | Verbeke, Wouter, David Martens, and Bart Baesens. (2014) [126] |
| Telecom. | Churn Prevention, Customer Profitability | Bahnsen, Alejandro Correa, Djamila Aouada, and Björn Ottersten. (2015) [66] | Telecom. | Churn Prevention | Gerpott et al. (2001) [134] |
| Telecom. | Churn Prevention | Seo et al. (2008) [135] | Telecom. | Churn Prevention | Pendharkar (2009) [136] |
| Telecom. | Churn Prevention | Kim and Yoon (2004) [139] | Telecom. | Churn Prevention | Amin, Adnan, et al. (2016) [144] |
| Telecom. | Churn Prevention | Gui, Chun. (2017) [145] | Telecom. | Churn Prevention | Xia, Guo-en, and Wei-dong Jin. (2008) [39] |
| Telecom. | Churn Prevention | Kirui, Clement, et al. (2013) [116] | | | |

tomer’s behavioral patterns from vast amount of data by layer-wise stacked neurons structure. Therefore, given minute timestamps and abundant observations, applying this data to deep learning algorithms for the generation of latent features is expected to produce better performance than conventional churn prediction models.

This is because as the log data in these days is collected for a longer period and deep learning algorithm get an advantage to catch customers’ latent status compared with older algorithms. In other words, the reason deep learning algorithms are receiving spotlight today is due to the vast amount of data used in modern churn predictions, and its ability to capture minute changes. As mentioned earlier in the text, traditional churn prediction algorithms including statistics methods are still actively used today. This is due to variations in which churn prediction model has the best performance depending on the data format. Churn prediction models using deep learning is a new solution with a good structure for predicting modern churn datasets. Therefore, to solve the problem at hand, readers will need to understand the format of the churn dataset and apply a suitable algorithm to solve the churn prediction problem.

Furthermore, we also outlined a performance evaluation method for comparing the various churn prediction algorithms used from the past to the present. Most churn prediction models are related to customer relation management. For example, there may be performance differences depending on whether the churn prediction model is robust against false positives or false negatives. According to the research of this paper, many articles use AUC as a performance measurement method aside from standard precision. In general, as there are fewer churn customers than non-churn customers, a performance specific method focused on churn customers will be needed. The ROC curve is a graph of the rate at which the model correctly predicts churn customers and the rate at which residual customers are predicted to
| Model Summary                      | Prediction Algorithms     | Best           | Publishing Information                      | Model Summary                      | Prediction Algorithms     | Best           | Publishing Information                      |
|------------------------------------|---------------------------|----------------|---------------------------------------------|------------------------------------|---------------------------|----------------|---------------------------------------------|
| Traditional Machine Learning      | AdaBoost, DT, NB          | AdaBoost       | Kawale, Jaya, Aditya Pal, and Jaideep Srivastava. (2009) [78] | Traditional Machine Learning      | Apriori Algorithm, SVM, NB | SVM            | Morik, Katharina, and Hanna Köpcke. (2004) [98] |
| Traditional Machine Learning      | Association Rule          | Association Rule | Chiang, Ding-An, et al. (2003) [35]            | Traditional Machine Learning      | Bagging, NN, Logit, ADT, BN, RF, SVM | NN             | Verbraken, Thomas, Wouter Verbeke, and Bart Baesens. (2012) [55] |
| Traditional Machine Learning      | Bagging, RF, RSM, CART, Rotation Forest | Rotation Forest | De Bock, Koen W., and Dirk Van den Poel. (2011) [103] | Traditional Machine Learning      | Boosting, Bagging, Logit | Boosting       | Lemmens, Aurélie, and Christophe Croux. (2006) [119] |
| Traditional Machine Learning      | Boosting, Logit, DT, NN, RF | Boosting       | Clemente, M., V. Giner-Bosch, and S. San Matias. (2010) [97] | Traditional Machine Learning      | Boosting, Logit, SVM | Boosting       | Tamaddoni, Ali, Stanislav Stakhovyich, and Michael Ewing. (2016) [102] |
| Traditional Machine Learning      | CART                       | CART           | Tamaddoni Jahromi, Ali, et al. (2010) [20]    | Traditional Machine Learning      | DCES, LR, ANN, SVM, NB | DCES           | Baumann, Annika, et al. (2015) [109]           |
| Traditional Machine Learning      | DT                         | DT             | Wei, Chih-Ping, and I-Tang Chu. (2002) [41]   | Traditional Machine Learning      | DT                         | DT             | Richter, Yossi, Elad Yom-Tov, and Noam Slonim. (2010) [53] |
| Traditional Machine Learning      | DT, GAM, RF + GAM ensemble | DT             | Coussette, Kristof, and Koen W. De Bock. (2013) [88] | Traditional Machine Learning      | DT, Logit, KNN, SVM, K-means Clustering | DT             | Huang, Ying, and Tahar Kechadi. (2013) [120] |
| Traditional Machine Learning      | DT, Logit, SVM             | SVM, DT        | Xie, Hanting, et al. (2015) [127]            | Traditional Machine Learning      | DT, Logit        | DT             | Dalvi, Preeti K., et al. (2016) [107]            |
| Traditional Machine Learning      | DT, LR                     | DT             | Dahiya, Kiran, and Surbhi Bhatia. (2015) [65] | Traditional Machine Learning      | DT, LR, NN, NB | DT             | Hadiji, Fabian, et al. (2014) [42]            |
| Traditional Machine Learning      | DT, LR, RF                 | DT             | Bahnusen, Alejandro Correa, Djamila Aouada, and Björn Ottersten. (2015) [66] | Traditional Machine Learning      | DT, NB, BN | DT             | Kirui, Clement, et al. (2013) [116]            |

*Abbreviations were defined in Appendix B-B.*
be churn customers. Therefore, it is a performance measurement method that focuses on the prediction of churn customers. In this study, we comprehensively compared the churn prediction problems. This paper helps to find a method that meets the needs of researchers among various churn prediction algorithms. Furthermore, this paper is expected to

| TABLE 4. (Continued.) Churn prediction models. |
|-----------------------------------------------|
| **Traditional Machine Learning** | **DT, NN, LR, CART** | **DT** | **Hadden, John, et al. (2006) [117]** | **Traditional Machine Learning** | **GAM, LR** | **GAM** | **Coussen, Kristof, Dries F. Benoit, and Dirk Van den Poel. (2010) [101]** |
| **Traditional Machine Learning** | **K-means, NMF, PCA, SIVM, AA** | **AA** | **Bindewald, Jason M., Gilbert L. Peterson, and Michael E. Miller. (2016) [70]** | **Traditional Machine Learning** | **KNN** | **KNN** | **Castro, Emiliano G., and Marcos SG Tsuzuki. (2015) [86]** |
| **Traditional Machine Learning** | **KNN** | **KNN** | **Borbor, Zoheb, et al. (2011) [47]** | **Traditional Machine Learning** | **KNN, DT, RF, Logit** | **KNN** | **Bastiaan van der Palen. (2017) [76]** |
| **Traditional Machine Learning** | **Logit** | **Logit** | **Shores, Kenneth B., et al. (2014) [84]** | **Traditional Machine Learning** | **NN, SVM, DT** | **SVM + AdaBoost** | **Vafeiadis, Thanasis, et al. (2015) [8]** |
| **Traditional Machine Learning** | **Logit, LR, NB, MLP, SVM** | **Logit, SVM** | **Huang, Bingquan, Mohand Tahar Kechadi, and Brian Buckley. (2012) [121]** | **Traditional Machine Learning** | **Logit, MLN** | **MLN** | **Dierkes, Torsten, Martin Bichler, and Ramayya Krishnan. (2011) [123]** |
| **Traditional Machine Learning** | **Logit, RF, PCA, XGB** | **XGB** | **Ge, Yizhe, et al. (2017) [108]** | **Traditional Machine Learning** | **NB, DT, KNN, GB, AdaBoost, Logit, RF, HMM** | **HMM** | **Tamassia, Marco, et al. (2016) [75]** |
| **Traditional Machine Learning** | **Logit, DT** | **Logit, DT** | **Ali, Özden Gür, and Umut Artürk. (2014) [94]** | **Traditional Machine Learning** | **Logit, DT** | **Logit** | **Risselada, Hans, Peter C. Verhoef, and Tammo HA Bijmolt. (2010) [100]** |
| **Traditional Machine Learning** | **Logit, SVM, C45** | **SVM** | **Moeversoms, Julie, and David Martens. (2015) [51]** | **Traditional Machine Learning** | **LR, DT** | **LR** | **Nie, Guangli, et al. (2011) [93]** |
| **Traditional Machine Learning** | **LR, DT** | **DT** | **Milošević, Miloš, Nenad Živić, and Igor Andjelković. (2017) [44]** | **Traditional Machine Learning** | **LR, DT** | **LR** | **Radosavljevik, Dejan, Peter van der Putten, and Kim Kyllesbech Larsen. (2010) [68]** |

*Abbreviations were defined in Appendix B-B.*
| Table 4. (Continued.) Churn prediction models. |
|---------------------------------------------|
| Traditional Machine Learning | LR, DT, NN, AdaBoost, CSDT | AdaBoost, CSDT | GLady, Nicolas, Bart Baesens, and Christophe Croux. (2009) [38] | Traditional Machine Learning | LR, DT, NN, BN | DT | Neslin, Scott A., et al. (2006) [54] |
| Traditional Machine Learning | LR, RF, GB | RF | Burez, Jonathan, and Dirk Van den Poel. (2009) [95] | Traditional Machine Learning | LR, SVM, DT, RF | LR | Yang, Wanshan, et al. (2019) [43] |
| Traditional Machine Learning | NB, Logit, TAN, Max-Min Hill Climbing | Logit | Verbraken, Thomas, Wouter Verbeke, and Bart Baesens. (2014) [4] | Traditional Machine Learning | NB, DT, NN | NN | Nozhnin, Dmitry. (2012) [74] |
| Traditional Machine Learning | NB, Logit, SVM, DT, RF, KNN | RF | Dror, Gideon, et al. (2012) [57] | Traditional Machine Learning | NB | NB | Nath, Shyam V., and Ravi S. Behara. (2003) [118] |
| Traditional Machine Learning | DT, Logit | Logit | Lim, Tong-Ming, and Angela Siew Hoong Lee. (2017) [152] | Traditional Machine Learning | NN, DMEL | DMEL | Au, Wai-Ho, Keith CC Chan, and Xin Yao. (2003) [34] |
| Traditional Machine Learning | NN, DT, SVM, ESVM | ESVM | Yu, Xiaobing, et al. (2011) [106] | Traditional Machine Learning | NN, K-means Clustering + NN, SOM + NN | NN + NN | Hudaib, Amjad, et al. (2015) [112] |
| Traditional Machine Learning | NN, Logit, DT, SVM | NN | Runge, Julian, et al. (2014) [45] | Traditional Machine Learning | NN, Logit | NN | Mozer, Michael C., et al. (2000) [16] |
| Traditional Machine Learning | NN, LR, RF | RF | Buckinx, Wouter, and Dirk Van den Poel. (2005) [96] | Traditional Machine Learning | NN, SOM | NN | Tsai, Chih-Fong, and Yu-Hsin Lu. (2009) [111] |
| Traditional Machine Learning | SVM | SVM | Hur, Yeon, and Sehun Lim. (2005) [99] | Traditional Machine Learning | SVM, MLP, DT, NN | NN | Keramati, Abbas, et al. (2014) [125] |
| Traditional Machine Learning | SVM, RF, NB | SVM | Saradhi, V. Vijaya, and Girish Keshav Palshikar. (2011) [50] | Traditional Machine Learning | SVM, RF, Logit | RF | Cousseng, Kristof, and Dirk Van den Poel. (2008) [67] |
| Traditional Machine Learning | RF, DT, SVM | RF | Sifa, Rafat, et al. (2015) [52] | Traditional Machine Learning | RF, KNN | RF | Idris, Adnan, Muhammad Rizwan, and Asifullah Khan. (2012) [122] |

*Abbreviations were defined in Appendix B-B.
| TABLE 4. (Continued.) Churn prediction models. |
|---------------------------------------------|
| **Traditional Machine Learning** | **XGBoost, NN, DT, SVM** | **XGBoost** | **Xie, Yaya, et al. (2009) [91]** | **Traditional Machine Learning** | **RF, SVM, LR, DT, MLP** | **RF** | **Anil Kumar, Dudyala, and Vadlamani Ravi. (2008) [92]** |
| **Traditional Machine Learning** | **RF, XGBoost, GBM** | **RF** | **Lee, Bunjo, et al. (2018) [17]** | **Traditional Machine Learning** | **RF, XGBoost, GBM** | **RF** | **Lee, Bunjo. (2019) [82]** |
| **Traditional Machine Learning** | **SVM** | **SVM** | **Ngomnang, Blaise, Emmanuel Viennet, and Maurice Tchuenté. (2012) [105]** | **Deep Learning** | **DNN + LSTM** | **DNN + LSTM** | **Guitart, Anna, Pei Pei Chen, and África Periáñez. (2019) [80]** |
| **Deep Learning** | **Logit, GB, RF, LSTM** | **GB** | **Kim, Seungwook, et al. (2017) [69]** | **Deep Learning** | **DSM, CNN, LSTM, SGD, GB, RF** | **DSM** | **Zhang, Rong, et al. (2017) [11]** |
| **Deep Learning** | **CNN** | **CNN** | **Mishra, Abinash, and U. Srinivasulu Reddy. (2017) [129]** | **Deep Learning** | **CNN** | **CNN** | **Umapaparvathy, V., and K. Iyakutti. (2017) [130]** |
| **Deep Learning** | **LSTM** | **LSTM** | **Evermann, Joerg, Jana-Rebecca Rehse, and Peter Fettke. (2017) [18]** | **Deep Learning** | **RF, NN, LSTM, Hidden State LSTM** | **Hidden State LSTM** | **Kristensen, Jeppe Theiss, and Paolo Burelli. (2019) [79]** |
| **Deep Learning** | **CNN + Autoencoders** | **CNN + Autoencoders** | **Wangperawong, Arit, et al. (2016) [115]** | **Graph Theory** | **Network Analysis (Centrality), MLP** | **Network Analysis** | **Kim, Kyoungok, Chi-Hyuk Jun, and Jaewook Lee. (2014) [124]** |
| **Graph Theory** | **Graph (CDRN, NLB, SPA RC, WVRN), DT, Bagging, RF, BN, Logit** | **Graph** | **Verbeke, Wouter, David Martens, and Bart Baesens. (2014) [126]** | **Graph Theory** | **Graph Energy distribution** | **Graph Energy distribution** | **Dasgupta, Koustuv, et al. (2008) [110]** |

*Abbreviations were defined in Appendix B-B.
be used to improve services and build better churn analysis models.

VIII. LIMITATIONS AND ISSUES FOR FURTHER RESEARCH

Churn studies on different fields are undoubtedly helpful in grasping the comprehensive view of churn and exploring various features to apply them to the churn models. However, as each study uses different sizes and types of features in the data, the set of studies provided in this paper has a limitation in comparing a common performance. Accordingly, although researchers may be able to discern whether if their constructed model is used widely through our study, they would not be able to determine which model is suitable and has the best performance for their study. Thus, in the future study, we intend to combine the feature engineering of fields introduced in this paper with the open churn datasets and construct various churn prediction models, including a deep learning model, and then conduct experiments on comparing and evaluating the suitability of each model.

APPENDIX A
CHURN ANALYSIS IN VARIOUS BUSINESS FIELD
See Table 3.

APPENDIX B
CHURN PREDICTION MODELS
A. CHURN PREDICTION MODELS
See Table 4.

**TABLE 4. (Continued.) Churn prediction models.**

| Interview | User Behavior Interview | - | Ben Lewis-Evans. (2012) [71] |
|-----------|-------------------------|---|----------------------------|
| Statistics | Cox Regression (Survival Analysis) | Survival Analysis | Bertens, Paul, Anna Guitart, and África Periáñez. (2017) [73] |
| Statistics | Survival Analysis | Survival Analysis | Statistics |
| Statistics | Survival Analysis | Bauckhage, Christian, et al. (2012) [83] |
| Statistics | Survival Analysis | Statistics |
| Statistics | Survival Analysis | Periáñez, África, et al. (2016) [14] |
| Statistics | Survival Analysis | Statistics |
| Statistics | Survival Analysis | Sifa, Rafet, Christian Bauckhage, and Anders Drachen. (2014) [56] |
| Statistics | Survival Analysis | Statistics |
| Statistics | Survival Analysis | Larivière, Bart, and Dirk Van den Poel. (2004) [36] |
| Statistics | Survival Analysis | Statistics |
| Statistics | Survival Analysis | Dechant, Andrea, Martin Spann, and Jan U. Becker. (2019) [46] |
| Statistics | Survival Analysis | Statistics |
| Statistics | Status Diagram with Hypothesis Test | Status Diagram with Hypothesis Test | Ahn, Jae-Hyeon, Sang-Pil Han, and Yung-Seop Lee. (2006) [33] |

*Abbreviations were defined in Appendix B-B.
Table 5. List of abbreviations.

| Abbreviation | Definition                                      | Abbreviation | Definition                                      |
|--------------|------------------------------------------------|--------------|------------------------------------------------|
| PCA          | Principal Component Analysis                   | NMF          | Non-negative Matrix Factorization              |
| NB           | Naive Bayes                                     | RF           | Random Forest                                  |
| SF           | Similarity Forest                               | DT           | Decision Tree                                  |
| GAM          | Generalized Additive Model                      | ADT          | Alternating Decision Tree                      |
| CSDT         | Cost Sensitive Decision Tree                    | CART         | Classification and Regression Tree             |
| TAN          | Tree Augmented Naive bayes                      | SVM          | Support Vector Machine                         |
| ESVM         | Evolutionary Support Vector Machine             | DMEEL        | Data Mining by Evolutionary Learning           |
| SOM          | Self-Organizing Maps                            | RSM          | Random Subspace Method                         |
| Logit        | Logistic Regression                             | LR           | Linear Regression                              |
| KNN          | K-Nearest Neighbors                             | DCES         | Decision Centric Ensemble Selection           |
| GB           | Gradient Boosting                               | XGB          | eXtreme Gradient Boosting                      |
| HMM          | Hidden Markov Model                             | MLP          | Multilayer Perceptron                         |
| BN           | Bayesian Network                                | MLN          | Markov Logic Network                          |
| SGD          | Stochastic Gradient Descent                     | DNN          | Deep Neural Network                           |
| CNN          | Convolutional Neural Network                    | LSTM         | Long Short-Term Memory model                  |
| DSM          | Deep Shallow mixture Model                      | DTW          | Dynamic Time Wrapping                         |
| AA           | Archetypal Analysis                             | ARIMA        | Autoregressive Integrated Moving Average      |
| NLB          | Network-only Link Based classifier              | CDRN         | Class-Distribution Relational Neighbor classifier |
| WVRN         | Weighted-Vote Relational Neighbor               | SPA-RC       | SSpatial Association Rule-based Classifier     |

B. LIST OF ABBREVIATIONS

See Table 5.

REFERENCES

[1] M. Chandar and P. A. L. Krishna, “Modeling churn behavior of bank customers using predictive data mining techniques,” in Proc. Nat. Conf. Soft Comput. Techn. Eng. Appl. (SCT), 2006, pp. 24–26.

[2] A. Parvatiyar and J. N. Sheth, “Customer relationship management: Emerging practice, process, and discipline,” J. Econ. Social Res., vol. 3, no. 2, 2001.

[3] T. Fields, Mobile & Social Game Design: Monetization Methods and Mechanics. Boca Raton, FL, USA: CRC Press, 2014.

[4] T. Verbraken, W. Verbeke, and B. Baesens, “Profit optimizing customer churn prediction with Bayesian network classifiers,” Intell. Data Anal., vol. 18, no. 1, pp. 3–24, Jan. 2014.

[5] Investopedia. Accessed: May 4, 2020. [Online]. Available: https://www.investopedia.com/terms/c/churnrate.asp

[6] A. M. Almana and M. S. A. R. Alzahrani, “A survey on data mining techniques in customer churn analysis for telecom industry,” Int. J. Eng. Res. Appl., vol. 4, no. 5, pp. 165–171, 2014.

[7] A. Ahmed and D. M. Lienen, “A review and analysis of churn prediction methods for customer retention in telecom industries,” in Proc. 4th Int. Conf. Adv. Comput. Commun. Syst. (ICACCS), Jan. 2017, pp. 1–7.

[8] T. Vafeiadis, K. I. Diamantas, G. Sarigiannidis, and K. C. Chatzisavvas, “A comparison of machine learning techniques for customer churn prediction,” Simul. Model. Pract. Theory, vol. 55, pp. 1–9, Jun. 2015.

[9] M. Ahmed, H. Afzal, A. Majeed, and B. Khan, “A survey of evolution in approaches to churn as a business problem: A survey,” Knowl. Inf. Syst., vol. 51, no. 3, pp. 719–774, Jun. 2017.

[10] Y. Chen, X. Xie, S.-D. Lin, and A. Chia, “WSDM cup 2018: Music recommendation and churn prediction,” in Proc. 11th ACM Int. Conf. Web Search Data Mining (WSDM), 2018, pp. 8–9.

[11] M. C. Mozer, R. Wolniewicz, D. B. Grimes, E. Johnson, and H. Kaushansky, “Predicting subscriber dissatisfaction and improving retention in the wireless telecommunications industry,” IEEE Trans. Neural Netw., vol. 11, no. 3, pp. 690–696, May 2000.

[12] E. Lee, B. Kim, S. Kang, B. Kang, Y. Jang, and H. K. Kim, “Profit optimizing churn prediction for long-term loyal customers in online games,” IEEE Trans. Games, vol. 12, no. 1, pp. 41–53, Mar. 2020.

[13] J. Evermann, J.-R. Rehse, and P. Fettke, “Predicting process behaviour using deep learning,” Decis. Support Syst., vol. 100, pp. 129–140, Aug. 2017.

[14] M. Hossouna, A. Tarhini, T. Elyas, and M. S. A. Trab, “Customer churn in mobile markets: A comparison of techniques,” Int. Bus. Res., vol. 8, no. 6, pp. 1–14, May 2015.

[15] A. Perianez, A. Saas, A. Guitart, and C. Magne, “Churn prediction in mobile social games: Towards a complete assessment using survival ensembles,” in Proc. IEEE Int. Conf. Data Sci. Adv. Analytics (DSAA), Oct. 2016, pp. 564–573.

[16] Z. He, X. Xu, J. Z. Huang, and S. Deng, “Mining class outliers: Concepts, techniques and applications in CRM,” Expert Syst. Appl., vol. 27, no. 4, pp. 681–697, Nov. 2004.
J. Ahn
VOLUME 8, 2020
220837

[94] Ö. Gür Ali and U. Aritürk, “Dynamic churn prediction framework with

[93] G. Nie, W. Rowe, L. Zhang, Y. Tian, and Y. Shi, “Credit card churn

[92] L. Butgereit, “Work towards using micro-services to build a data pipeline

[91] K. Coussement and K. W. De Bock, “Customer churn prediction in the

[90] H. Wolters, J. Baer, and G. Keswani, “Method to detect and score churn

[89] T. Debeauvais, B. Nardi, D. J. Schiano, N. Ducheneaut, and N. Yee,

[88] K. B. Shores, Y. He, K. L. Swanenburg, R. Kraut, and J. Riedl, “The iden-

[87] A. Guitart, P. Chen, and Á. Periáñez, “The winning solution to the

[86] T. Fields and B. Cotton, [85] M. Tamassia, W. Raffe, R. Sifa, A. Drachen, F. Zambetta, and

[84] P. Bertens, A. Guitart, and A. Guitart, “Games and big data: A scalable

[83] J. Kawale, A. Pal, and J. Srivastava, “Churn prediction in MMORPGs:

[82] A. Guitart, P. Chen, and Á. Periáñez, “The winning solution to the

[81] T. Fields and B. Cotton, [80] M. Tamassia, W. Raffe, R. Sifa, A. Drachen, F. Zambetta, and

[79] J. Kawale, A. Pal, and J. Srivastava, “Churn prediction in MMORPGs:

[78] J. Kawale, A. Pal, and J. Srivastava, “Churn prediction in MMORPGs:

[77] J. Kawale, A. Pal, and J. Srivastava, “Churn prediction in MMORPGs:

[76] M. Chen, “Music streaming service prediction with MapReduce-based

[75] T. Fields and B. Cotton,

[74] T. Fields and B. Cotton,

[73] M. Hitchens, “Predicting player churn in destiny: A hidden Markov

[72] A. Guitart, P. Chen, and Á. Periáñez, “The winning solution to the

[71] J. Ahn

[70] J. Ahn

[69] J. Ahn

[68] J. Ahn

[67] J. Ahn

[66] J. Ahn

[65] J. Ahn

[64] J. Ahn

[63] J. Ahn

[62] J. Ahn

[61] J. Ahn

[60] J. Ahn

[59] J. Ahn

[58] J. Ahn

[57] J. Ahn

[56] J. Ahn

[55] J. Ahn

[54] J. Ahn

[53] J. Ahn

[52] J. Ahn

[51] J. Ahn

[50] J. Ahn

[49] J. Ahn

[48] J. Ahn

[47] J. Ahn

[46] J. Ahn

[45] J. Ahn

[44] J. Ahn

[43] J. Ahn

[42] J. Ahn

[41] J. Ahn

[40] J. Ahn

[39] J. Ahn

[38] J. Ahn

[37] J. Ahn

[36] J. Ahn

[35] J. Ahn

[34] J. Ahn

[33] J. Ahn

[32] J. Ahn

[31] J. Ahn

[30] J. Ahn

[29] J. Ahn

[28] J. Ahn

[27] J. Ahn

[26] J. Ahn

[25] J. Ahn

[24] J. Ahn

[23] J. Ahn

[22] J. Ahn

[21] J. Ahn

[20] J. Ahn

[19] J. Ahn

[18] J. Ahn

[17] J. Ahn

[16] J. Ahn

[15] J. Ahn

[14] J. Ahn

[13] J. Ahn

[12] J. Ahn

[11] J. Ahn

[10] J. Ahn

[9] J. Ahn

[8] J. Ahn

[7] J. Ahn

[6] J. Ahn

[5] J. Ahn

[4] J. Ahn

[3] J. Ahn

[2] J. Ahn

[1] J. Ahn

[0] J. Ahn

[220837]
A. D. Athanasopoulos, “Customer satisfaction cues to support market segmentation and explain switching behavior,” J. Bus. Res., vol. 47, no. 3, pp. 191–207, Mar. 2000.

H.-S. Kim and C.-H. Yoon, “Determinants of subscriber churn and customer loyalty in the Korean mobile telephony market,” Telecommun. Policy, vol. 28, nos. 9–10, pp. 751–765, Oct. 2004.

J. R. Miglautsch, “Thoughts on RFM scoring,” J. Database Marketing Customer Strategy Manage., vol. 8, no. 1, pp. 67–72, Aug. 2000.

L. O’Brien and C. Jones, “Do rewards really create loyalty?” Long Range Planning, vol. 28, no. 4, p. 130, 1995.

D. A. Schweidel and G. Knox, “Incorporating direct marketing activity into latent attrition models,” Marketing Sci., vol. 32, no. 3, pp. 471–487, May 2013.

L. Zhu and N. Laptev, “Deep and confident prediction for time series at uber,” in Proc. IEEE Int. Conf. Data Mining Workshops (ICDMW), Nov. 2017, pp. 103–110.

A. Amin, S. Anwar, A. Adnan, M. Nawaz, N. Howard, J. Qadir, A. Hawalah, and A. Hussain, “Comparing oversampling techniques to handle the class imbalance problem: A customer churn prediction case study,” IEEE Access, vol. 4, pp. 7940–7957, 2016.

C. Gui, “Analysis of imbalanced data set problem: The case of churn prediction for telecommunication,” Artif. Intell. Res., vol. 6, no. 2, p. 93, Jun. 2017.

W. J. Reinartz and V. Kumar, “On the profitability of long-life customers in a noncontractual setting: An empirical investigation and implications for marketing,” J. Marketing, vol. 64, no. 4, pp. 17–35, Oct. 2000.

P. S. Fader, B. G. S. Hardie, and J. Shang, “Customer-base analysis in a discrete-time noncontractual setting,” Marketing Sci., vol. 29, no. 6, pp. 1086–1108, Nov. 2010.

P. S. Fader and B. G. H. K. L. Lee, “Counting your customers’ the easy way: An alternative to the Pareto/NBD model,” Marketing Sci., vol. 24, no. 2, pp. 275–284, 2005.

P. S. Fader and B. G. S. Hardie, “Probability models for customer-base analysis,” J. Interact. Marketing, vol. 23, no. 1, pp. 61–69, Feb. 2009.

P. S. Fader and B. G. S. Hardie, “How to project customer retention,” J. Interact. Marketing, vol. 21, no. 1, pp. 76–90, Jan. 2007.

D. G. Morrison and D. C. Schmittlein, “Generalizing the NBD model for customer purchases: What are the implications and is it worth the effort?” J. Bus. Econ. Statist., vol. 6, no. 2, pp. 145–159, Apr. 1988.

T.-M. Lim and A. S. H. Lee, “Loyalty card membership challenge: A study on membership churn and their spending behaviour,” Arch. Bus. Res., vol. 5, no. 6, 2017. [Online]. Available: http://j116.203.177.230/index.php/ABR/article/view/3054, doi: 10.14738/abr.56.3054.
DOYOUNG KIM received the B.A. degree in economics from Yonsei University in 2019. He is currently an Assistant Data Scientist with Intelligence Laboratories, Nexon Korea.

HYUKGEUN CHOI received the B.S. degree in computer science from Seoul National University in 2009, the M.S.E.E. degree in 2017, and the Ph.D. degree in electrical engineering and computer science from Seoul National University in 2019. He joined as an AI Engineer with Intelligence Labs, Nexon Korea.

SHINJIN KANG (Member, IEEE) received the M.S. degree from the Department of Computer Science and Engineering, Korea University, in 2003, and the Ph.D. degree in computer science and engineering from Korea University in 2011. He joined Sony Computer Entertainment Korea as a Video Game Developer. Since 2006, he has been a Lead Game Designer with NCsoft Korea. He is currently a Professor with the School of Games, Hongik University.

* * *