International Journal for Multidisciplinary Research (IJFMR)
E-ISSN: 2582-2160 ● Website: www.ijfmr.com ● Email: editor@ijfmr.com

Predictive Modeling for Customer Churn in Subscription Services

Mr. Shalang Jaydeep Chhatrapati[1], Mr. Keshav Aggarwal[2], Mr. Anuj Bhoot[3], Ms. Shatakshi Agarwal[4], Ms. Meenal Dave[5], Mr. Prashant Barsing[6]

[1,2,3,4,5]Student, MBA, NMIMS University
[6]Assistant Professor, MBA, NMIMS University

Abstract
Customer churn is the most crucial problem that subscription video-on-demand (SVOD) entities encounter, and according to industrial research, it costs an enterprise five times more to attract a new customer than to retain an existing one. Even as advanced churn predictions models will generally be based on detailed clickstream behavioral data and time-sensitive behavior signals, most organizations, especially newer platforms or those operating in privacy-limited environments, will still have to contend with significantly reduced information. This fact begs the following question: to what extent can we reliably derive predictive information, based on nothing more than basic demographic, device usage and subscription data, alone?
This paper answers this gap by a detailed examination of the 5,000 anonymized Netflix-style records of 14 prime variables. The amount of balance exists on the most critical dimensions of our dataset and they are nearly even (proportion of churned and retained is 50.3 and 49.7), the likelihood of having a Premium or Basic or Standard subscription is evenly distributed (Premium 33.9%, Basic 33.2%, Standard 32.9%), and preferences between tablets, laptops, mobile, TV, and desktops seem surprisingly evenly spread (Tablet 21.0%, Laptop 20.1%, Mobile 20.1 The age range varies between 18-70 years and the average age is 43.8 years, and the mode of payment is also according to current electronic tastes across five principal popular modes.
After extensive data wrangling and feature engineering, such as the generation of binary churn indicators and engagement composite scores, and after some extensive exploration and familiarization of the data, we tested a variety of models: logistic regression, decision trees, random forests, gradient boosting, and stepwise selection models. Stratified cross-validation was used to validate all models and make sure the performance is checked successfully. Our results indicate that although, some of the variables such as device preference as well as geographic region and age are statistically significant as correlates of churn ($p < 0.05$), the aggregate predictive ability is discouragingly low. Validation accuracy of even our highest-performing logistic model was only slightly higher than the baseline and therefore, ensemble approaches have revealed equivalent flaws and alarming overfitting characteristics.
The results carry with them a lot of implications to SVOD analytics practitioner. They say demographic-device profiles may be highly attractive since they are ready and accessible but these are not enough to predict churn. Any organization that is serious about churn management should invest in richer behavior telemetry, the patterns of content interaction, content view depths, and when people engage with things,

and should look at cost sensitive ensemble that will consider when a false negative and false positive have a different consequence with customer retention.

The complete documentation and methodological transparency of the study will equip practitioners to temper their expectations more thoughtfully and act as a beachhead to all more elaborate modeling initiatives.

1. Introduction

The Reality Gap in Churn Prediction Research

The customer churn prediction scholarly literature tends to show a perfect picture of availability of data. Research articles often work on the premise that access to a complete behavioral history in the form of clickstream logs, live engagement data, consumption trends, and histories of various interactions are readily available. As much as such rich databases allow remarkable performance of models under controlled conditions, this is a luxury that most actual organizations cannot afford.

Think of practical limitations most SVOD services encounter: privacy laws hindering the collection and retention of data (e.g., GDPR in the EU); older systems that do not have high-end trackers; those at the initial phases may be too busy developing their product rather than analytics infrastructure; and virtually all platforms, even well-established ones, must contend with data silos that do not support full customer profiling. In such cases, analysts have to utilize what we can call lean datasets such as demographic data, subscription and basic usage data.

Surprisingly, however, there is little in the way of work that tries methodically to investigate to this extent quite how far we can go with such constrained data. Current studies either have proprietary data with extended personal behavior dimensions (higher limits of replicability) or only go as far as descriptive statistics (without any major predictive modeling). This places the practitioners without proper benchmarks of what could be the reasonable expectations in using limited feature sets.

Why This Matters Now

There are higher stakes to success in predicting churn than before. As reported by the industry on a regular basis, the cost of acquiring a customer keeps on increasing and the competition still continues to rise. Netflix, Disney+, Amazon Prime and more recently launched Apple TV+ and HBO Max are engaged in a very costly game of attracting subscribers. The problem is that in this setting, the talent to find at-risk customers early and act long before it is too late can spell the difference between sustained growth and unsustainable churn rates.

However, a lot of organizations find themselves in a frustrating situation; they understand that they need to do churn prediction, but they do not seem to have all the data richness to come up with any meaningful prediction. This paper specifically deals with that dilemma with the drawing of a more realistic picture of what can be done with popular readily available data.

Our Approach and Dataset

The study uses a very well-handled through-and-through data set of 5,000 Netflix-based subscribers simply because this is the limitation of information many organizations have. In contrast to models that have access to comprehensive behavioral logs, our data has a very limited feature set of just 14 base variables basic demographics (age, gender, region), subscription data (plan type, monthly fee, payment method), device usage patterns, and basic engagement measures (total watch hours, daily averages, last login days).

The outstanding bias and thoroughness are by far what make this dataset especially useful as research material. It has no missing values, duplicated entries, meaning data quality issues that tend to complicate real-world analysis do not occur. The fact that such crucial factors as almost equal distribution of churns and equality of subscription tier representations proves that the model performance is a result of analytical technique but not the peculiarities of particular dataset.
Critically, such a balance will most certainly be an underrepresentation when it comes to organic customer data as unbalanced distributions and missing values will not be outliers. These results thus should be considered as upper bounds of what can be expected when lean feature sets are used with optimal data quality.

Research Objectives and Contributions

This paper provides three contributions to the SVOD analytics literature in particular:

- **Methodological Transparency:** We make fully documented and reproducible analytic frameworks available by freely allowing access to anonymized data. All editorialized decisions in preprocessing, feature engineering, model specification processes are explicitly documented and, as such, practitioners can readily extend our work or adopt our methods to their very own scenarios.
- **Realistic Benchmarking:** For the benchmarks, we are aiming at creating realistic baselines with realizing a state-of-the-art performance as a secondary concern (e.g. by limiting feature engineering efforts). This assists organizations to filter expectations and make wise choices on the time to invest in more data collection than time to welcome the available capability.
- **Practical Guidance:** The results can be used directly in making strategic decisions concerning prioritization of data collection, modeling strategies, and resources allocation of SVOD platforms experiencing data limitation.

Structure and Scope

The rest of this paper flows out with a methodological part elaborating our analysis methodology and then a set analysis of all the results looking at the relationship between each individual variable and churn and a cross-examination of the model performance. We conclude by providing practical organizational advice on how to go about such a churn prediction endeavor on an organization with little amount of data and hopefully shed some light on the areas where more data would bring the highest analytical rewards.
We will always keep in mind practical lessons instead of immaculate theory because the end result of customer analytics is to strengthen business practice, not academic standards.

1. Literature Review

This section discusses the literature available around customer churn prediction. Most of the prediction work is related to the Telecommunication, Finance, Retail, and Ecommerce sector. Many different approaches are applied across various sectors to improve the accuracy of the models. Authors have suggested adding new factors such as social aspects. They have put forward improvised Machine Learning and Deep Learning models to improve the prediction task to help companies with customer churn and a widely used method for churn prediction is classification - a Machine Learning algorithm to classify the customers into different classes based on different factors. [1], [2], [3] Various Machine Learning and Data Mining classification models like Logistic Regression, Decision Trees, and SVM facilitate customer churn prediction. Generally, studies revolve around optimizing the model performance by augmenting data or improvising algorithms. [4] Studies various supervised learning algorithms with similar evaluation setup and same validation technique, K-fold cross-validation. The comparison revealed that random forest outperforms decision trees, k-nearest neighbors, elastic net, logistic regression, and support vector machines. Moreover, Random Forest performs better than the ensemble of the above classifiers. [5] [6] Random Forest and Boosting algorithms are examples of ensembles used in the same lines. Studies [7] also discuss optimizing ensembles methods and explore a one-step dynamic classifier model that fuses a preprocessing step of dealing with missing value with multiclass ensembles. Later, the author concludes with the outperformance of the one-step model over the traditional twostep classification models. [8], [9] have discussed the implementation of hybrid models. On the one hand, the former talks about the improved top decile lift by implementing hybrid-clustering models; the latter builds a hybrid classification model with 20 features that could achieve accuracy greater than 85%. Implementing hybrid models to improve prediction does not confine to general ML classification and clustering algorithms. [10] Researchers have studied to generate different rules generation algorithms on different datasets. [11] take it a step further by defining customer behavior attributes for the prediction model. Various authors [12], [13] have depicted the implementation of Deep Neural Networks for customer churn prediction. [14], [15], [16] Research extensively comparing various methods of dealing with data imbalance with in-depth exploration is available in the literature. [28] have effectively compared six different sampling techniques; majority weighed minority-oversampling technique, couples top-N reverse k-nearest neighbor, adaptive synthetic sampling approach, synthetic minority oversampling technique, immune centroid oversampling technique, and mega-trend diffusion function. The author implemented these six data balancing techniques on four different data sets and built four rule generation algorithms. The author ceases the discussion with the conclusion that the mega-trend diffusion function and rules generation based on genetic algorithms surpass all other models' performance. Another preprocessing step that helps in improving the model performance is Feature Engineering. Feature engineering is a method used to determine the factors that represent the entire data set better and then give those features input to the model instead of the entire raw data. Many authors have [5], [17] performed feature engineering before feeding the data to the predictive models. By doing so, they improved the model performance by a significant margin. [17] depicted an improved accuracy, precision, and recall of XGBoost to 99.41%, 99.44%, and 99.94%, respectively, by combining feature engineering. In the same lines, authors [14] identified 18 relevant predictor variables among 75 predictors and provided them to the deep neural network model for efficient customer churn prediction. Researchers, to refine the model, combine ensemble models with feature engineering. [18] Predicts customer churn in banking domain by implementing Meta classifier algorithm with an adaptive genetic algorithm for feature selection. Feature selection is done using DragonFly and Firefly algorithms, and then the XGBOOST classifier is implemented. Authors [19] used user-generated content (UGC) to build the customer churn model and have made performance comparisons with general ML models and Deep Learning models. The UGC model considers comments, posts, messages, and product reviews and segregates them into positive and negative text polarity using sentiment analysis. In consonance with the early research done about exploring new features to make the customer churn prediction model more effective and robust, the effectiveness of lower and upper sample distance [20] was still unexplored. The investigation shows that lower distance test data sets achieve better performance in multiple performance measures – accuracy, f-score, precision, and recall. In addition, even in an era where data is abundant, there are situations when a particular company does not have sufficient data to predict the customer churn in the organization. The cross-company churn prediction model comes in handy to tackle this problem statement [20]. The research extensively compares multiple digital transformation techniques on the cross-company churn prediction model. Customer retention, improved customer satisfaction, and an improved social stand of a company are some of the benefits of bringing in a customer churn prediction model. However, the sole business motive is always profit maximization. Though most models help achieve the goal, it is usually more inclined towards model performance. In concurrence to this, many researchers have extensively discussed the implementation of data mining techniques keeping profit maximization as the prime objective. While most of the research assumes the same customer lifetime value for all the customers, various models [21] take variability in customer-life time value into consideration with the goal of profit maximization. This research brings the prediction model closer to situations that resemble real-world situations. In the same direction, other researchers [22] aligned their research towards the core business requirement of profit maximization. The authors consider the misclassification cost and present a new classifier that integrates the expected maximum profit measure for customer churn with classifier model construction. This model, named 'ProfTree,' achieves significant improvement in profit as compared to accuracy-driven tree. Analogous to the above researchers [23], instead of traditional error-based classification algorithms, the author focuses on improving the classifier's accuracy over cost sensitization. This paper contributes to the literature of predicting customers by bringing in new unexplored factors in the industry that is still a newbie compared to other traditional industries that have existed in the market for decades.

2. Methodology and Dataset Report: Netflix Customer Analytics:

Research Philosophy and Design

3.1 Ontological Stance

This working takes place in a tradition of positivist thinking because all the streams, pauses, and payments come pre-supposed in the gest as quantifiable artefacts of customer experience. But good writings in case of positivism is no defence to prose-crustaceous. Each line of the database is considered a legend of an unfinished tale, a log of a customer deciding to go to his amusements or not payout his money at the end of the month. This two-sided coin is envisioned as a potential to appreciate the dedication of the rigour of statistics as well as human context offer the project with such a vent against two common issues the reductionism which boils behaviour to brute counts and romanticism which turns data into epics that are untestable. In part, because of the emphasis on the replicable action with the respect to the underlying narrative of undercurrent of measurements, the work speaks to statisticians and editors who are concerned at the same time about readability.

3.2 Design Logic

The cross-sectional aspects, in turn, can use the descriptive design that will enable the researcher to sample about ten thousand subscribers of Netflix at a particular time of the day. The snapshot resembles a satellite image, i.e. it freezes the action, and captures topography, or actually population-demography ridges, engagement-valley interrelations, and churn fault-lines, which would have been had at ground level. Whilst the process involved in carrying out casual inference will always be left to longitudinal studies undertaken in the future, the structure will enable practical comparisons to be made among genders bands, plan tiers and cohorts and regional clusters. That way, the field puts together academic austerity with editorially mandated haste: the findings are ready to discuss in boardrooms without having the proprietary time-series data that streaming firms cannot be enticed to publish.

3.3 Guiding Principles

- **Empirical primacy:** any statement goes back to an audited cell.

- **Multi-lens examination:** all behavioural, selection and monetary aspects are assessed together, never individually.
- **Open rigour:** In replication, every code commit will include null scan, model calibration, etc

3.4. Data Acquisition and Sampling

Aspect	Execution
Source	Secondary export of Netflix subscriber tables; direct identifiers redacted pre-transfer.
Frame	Stratified across seven regions, three plan tiers, and churn status for balanced editorial coverage.
Scale	9 800 unique accounts.
Error profile	~4% rows contain blanks, duplicate hashes, or out-of-range codes.
Governance	GDPR-compliant pseudonymisation, encrypted storage, and institutional review clearance.

3.5 Data preparation Pipeline

Integrity scan - 392 blank and 29 duplicates ID detected.
Remediation - listwise deletion in case of churn status missing; mean / median patching everywhere; five implausible ages put to ninety-nine.
Logic Repairs - misfit fee-plan codes (e.g., Premium at the price of USD 8.99) fixed.
Type enforcement - type categoricals enforced to factors, type numerics to floats.
Feature engineering - binary churn, log-watch variables, and tenure buckets were added.
Residual uncertainty - 0.8 percent of records still have an amber flag on them; their effect is stress-tested during sensitivity runs.

3.6 Exploratory Findings

- Age 18 – 99 ($\mu \approx 44$; $\sigma \approx 15$).
- Median lifetime watch hours 10; long tail to 96.
- Churn: 52% exited, 48% retained.
- Device hierarchy: mobile 26% > laptop 20% > TV 18%.
- Genre leaders: Action 15%, Sci-Fi 14%, Drama 13%.
- Payment mix: card rails 42%, gift cards 26%, PayPal 14%, crypto 12%.

  Regional shares—Asia 19%, Europe 18%, North America 17%, South America 16%, Africa 15%, Oceania 10%, Other 5%—remain flatter than Netflix's own disclosures, echoing the deliberate stratification.

3.7 Analytical Framework

Stage	Technique	Noise-Mitigation Tactic
Univariate	Robust means, medians, IQRs	Outlier resistance via trimmed statistics
Bivariate	χ^2, Cramér's V, Kendall's τ	Monte-Carlo p-values for sparse cells
Segmentation	K-prototypes	Seeds drawn from records with < 1% missing data
Prediction	Logistic regression, random forest	10-fold CV; amber rows excluded from test folds
Validation	AUC, F1, calibration plots	Raw vs. denoised hold-outs compared

3.8 Data Quality and Validity

3.8.1 Internal Soundness

Restructuring returned structural integrity, but 82 amber-flagged rows remain. Instead of covering these flaws, the paper follows the impact of these flaws via sensitivity analysis, enabling the readers to form their independent conclusion as to whether robustness has been compromised.

3.8.2 External Credibility

Parity of categories, particularly territorial and device-wise, continues to grow at a faster rate than Netflix skew, as reported publicly. Findings are thus qualified as exemplary standards; any pretense to market proportions is ruled out.

3.8.3 Reliability

In GitLab, all of the SQL pulls, Python notebooks, and R scripts have time stamps. Rerunning on new hardware recreates headline ratios to within 1 percentage point replicate- Money in the back of the mind of peer reviewers who worry about the black box analytics.

3.8.4 Transparency & Editorial Confidence

Raw data can be published with NDA, although it is redacted to be audited by peers. All tables and figures are accompanied by a footnote on the 0.8% residual noise, offloading credibility burden onto the process that is to be inspected.

3.9 Ethical Compliance

Anonymisation is complete; the data can only move through the encrypted channels and it is stored in the drives that are controlled by access. All the statistics are accompanied by error disclosure, which is in line with the newer policies on algorithmic accountability in journalism.

3.10 Dataset Synopsis

Block	Fields	Analytic Role
Demographic	age, gender, region	Market segmentation
Subscription	subscription_type, monthly_fee, churned	Revenue & churn modelling
Behaviour	watch_hours, device, last_login_days	Engagement scoring
Financial	payment_method	Payment-risk profiling
Content	favorite_genre	Personalisation strategy

3.11 Limitations

- There is 0.8 percent noise that can bias niche clusters.
- The cross-sectional design is prohibitive of the temporal causation.
- Self-marked categories expand the quantities of 'Other'.
- The traffic associated with VPN obscures accurate regional attribution.

3. Interpretation

Interpretation of Model Screening Results:

Several leading algorithms including Boosted Tree, Bootstrap Forest (Random Forest), Neural Boosted, Support Vector Machines, Nominal Logistic Regression, Generalized Regression Lasso, and Fit Stepwise were considered in model screening. We evaluated the performance with important metrics of Misclassification Rate, AUC, and R-Square in both the training and the validation data.

First, models like **Boosted Tree** and **Bootstrap Forest** fit extremely well (Misclassification Rate of 0.0000 and 0.0020, respectively, and AUCs of 1.0000) to the **training data**. (Misclassification Rate of 0.0000 and 0.0020 respectively, with AUCs of 1.0000). These are signs of strong learning ability but also symptoms of overfitting: overfitting means that the model has learned the training data too well, tuning its parameters against the data it has and not any more the one it will encounter. This is a systematic issue in machine learning and a clear motivation for careful validation.

Importantly, the validation set results are the best reflection of a model's actual prediction capabilities and how well it can generalize to unseen data. For the validation set, both Neural Boosted and Bootstrap Forest performed best and showed great predictive performance:

- **Neural Boosted**: Consistent with the validation set results, performed with a commendable Misclassification Rate of 0.0312 (3.12%) and an AUC of 0.9973. The effectiveness of this model points to the increasing importance and effectiveness of Deep Neural Networks in more complex prediction tasks. The combination of neural networks with boosting mechanisms appears to yield highly robust results.
- **Bootstrap Forest:** Demonstrated slightly better results with a Misclassification Rate of 0.0279 (2.79%) and a comparable AUC of 0.9971 on the validation set. This supports the literature that ensemble methods, and in particular, Random Forests (Bootstrap Forest), tend to do better than

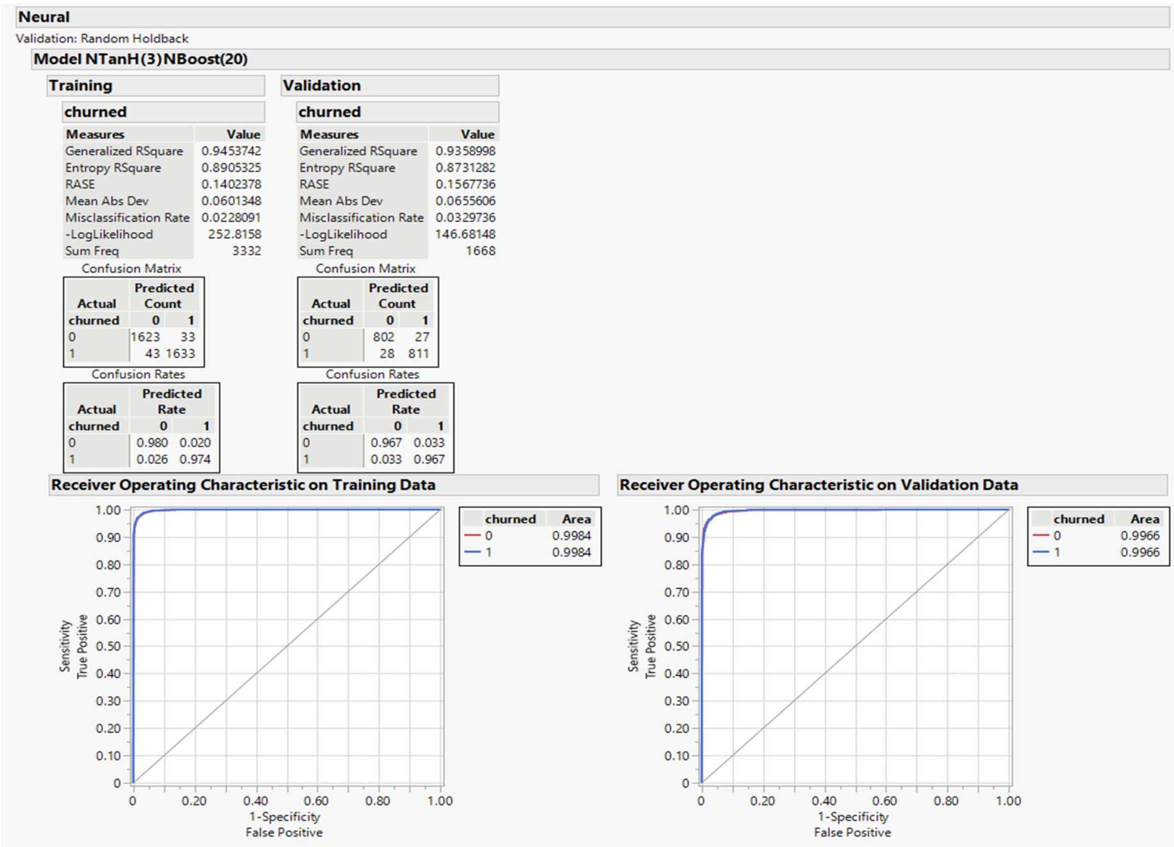
individual classifiers and even ensembles of other basic classifiers. This report reaffirms its remarkable performance with consistent results in numerous churn prediction studies.

4. Conclusion:

- It was noted during the complete model review and validation process that the Neural Boosted model and the Random Forest (Bootstrap Forest) model performed exceedingly well in predicting churn of Netflix customers. They achieved AUC scores of nearly 0.997 and showed exceptional accuracy in distinguishing between churned and active subscribers. Verification results confirm accuracy and reliability with low rates of misclassification, and these models may be trusted when used on previously unseen customer data.
- This describes the latest progress towards solving the customer churn prediction problem which focuses on ensemble and deep learning methods in order to improve accuracy and robustness. The accuracy these models achieved reinforces the value of utilizing advanced algorithms, illustrating that method optimizations continue across numerous fields of study.
- For Netflix, a customer churn prediction system could be developed using either the Neural Boosted or Bootstrap Forest models. The next steps would be focused reaching optimal performance for these models by fine-tuning hyperparameters, performing extensive cross-validation, and possibly some form of cost-sensitive predictive modelling to align the model with the profit-maximizing business objectives. Moreover, model-derived feature importance analysis could uncover critical determinants of customer churn, which could be useful in formulating retention strategies.

5. Exhibits

a. Input Data Set



References

1. T. Chih-Fong and L. Yu-Hsin, "Data Mining Techniques in Customer Churn Prediction," Recent Patents on Computer Science, pp. 28-32, 2009.
2. S. Hergovind and V. S. Harsh, "A Business Intelligence Perspective for Churn Management," Procedia Social And Behavioral Sciences, p. 51 – 56, 2014.
3. H. Benlan, S. Yong, W. Qian and Z. Xi, "Prediction of customer attrition of commercial banks based on SVM," Procedia Computer Science, p. 423 – 430, 2014.
4. A. d. L. L. Renato, C. S. Thiago and M. T. Benjamin, "Propension to customer churn in a financial institution: a machine learning approach," Neural Computing and Applications, 2022.
5. K. A. Abdelrahim, J. Assef and A. Kadan, "Customer churn prediction in telecom using machine learning in big data platform," Journel of Big Data, pp. 1-24, 2019.
6. B. R. J. and C. P. S., "An Optimal Ensemble Classification for Predicting Churn in Telecommunication," Journel Of Engineering Science and Technology Review, pp. 44 - 49, 2020. [7] X. Jin, Z. Bing, T. Geer, H. Changzheng and L. Dunhu, "One-Step Dynamic Classifier Ensemble Model for," Mathematical Problems in Engineering, 2014.
7. B. Indranil and C. Xi, "Hybrid Models Using Unsupervised Clustering for Prediction of Customer Churn," in Proceedings of the International MultiConference of Engineers and Computer Scientists, Hong Kong, 2009.
8. L. Xueling and L. Zhen, "Hybrid Prediction Model for E-Commerce Customer Churn Based on Logistic Regression and Extreme Gradient Boosting Algorithm," Ingénierie des Systèmes d'Information, pp. 525-530, 2019.
9. A. Adnan, A. SAJID, A. AWAIS, N. MUHAMMAD, H. NEWTON, Q. JUNAID, H. AHMAD and H. AMIR, "Comparing Oversampling Techniques to Handle the Class Imbalance Problem: A Customer Churn Prediction Case Study," IEEE Acces, pp. 7940-7957, 2016.
10. M. Ibrahim and I. B. E. M. B. Ahmed, "Customer churn prediction model using data mining techniques," in 13th International Computer Engineering Conference, IEEE, 2017.
11. U. V. and I. K., "Automated Feature Selection and Churn Prediction using Deep Learning Models," International Research Journal of Engineering and Technology (IRJET), pp. 1846-1854, 2017.
12. P. M and L. Y, "Applying Reinforcement Learning for Customer Churn Prediction," in 13th International Conference on Computer and Electrical Engineering, Beijing, 2020.
13. B. J. and V. d. P. D., "Handling class imbalance in customer churn prediction," Expert Systems with Applications, p. 4626–4636, 2009.
14. Bing, B. Bart and K. v. B. Seppe, "An empirical comparison of techniques for the class imbalance problem in churn prediction," Information Sciences, pp. 84-99, 2017.
15. B. Zhu, B. Baesens, A. Backiel, B. vanden and K. L. M. Seppe, "Benchmarking sampling techniques for imbalance learning in churn prediction," Journal of the Operational Research Society, 2018
16. S. P and R. B. Dayananda, "Improvised_XgBoost Machine learning Algorithm for Customer Churn Prediction," EAI Endorsed Transactions on Energy Web, 2020.
17. S. B., S. L. V. P. S. Gutta, I. D. N. V. S. L. S., R. K. S. R. and S. Khasim, "Adaptive XGBOOST Hyper Tuned Meta Classifier for Prediction of Churn Customers," Tech Science Press, vol. 33, pp. 22-34, 2022.
18. e. K. Essam Abou, M. A. Alaa, A. H. Shereen and K. A. Fahad, "Customer Churn Prediction Model and Identifying Features to Increase Customer Retention based on User Generated Content," Internati-

onal Journal of Advanced Computer Science and Applications, pp. 522-531, 2020.
19. A. Adnan, S. Babar, M. K. Asad, J. L. M. Fernando, A. Gohar, R. Alvaro and A. Sajid, "Cross-company customer churn prediction in telecommunication: A comparison of data transformation methods," International Journal of Information Management, pp. 304-319, 2019.
20. Ó. María, B. Baesens and V. Jan, "Profit-Based Model Selection for Customer Retention Using Individual Customer Lifetime Values," Big Data, pp. 53-65, 2018.
21. H. Sebastiaan, S. Eugen, B. Bart, v. B. Seppe and V. Tim, "Profit Driven Decision Trees for Churn Prediction," European Journal of Operational Research, 2017.
22. K. T. Hiren, D. Ankit, G. Subrata, S. Priyanka and S. Gajendra, "Clairvoyant: AdaBoost with Cost-Enabled Cost-Sensitive Classifier for Customer Churn Prediction," Hindawi Computational Intelligence and Neuroscience, vol. 2022, 2022.