

Customer Churn Analysis using Data-Driven Machine Learning Models

Name: Muhammad Burhan Aamir  
Module Code: COM64604

Module Name: Applied Machine Learning

Student Number: 2224726  
Email address: mba3crt@bolton.ac.uk

University Name: University Of Greater Manchester (UoGM), Bolton

*Abstract (short summary of report)*—This electronic document is a “live” template and already defines the components of your report [title, text, heads, etc.] in its style sheet. *\*CRITICAL: Do Not Use Symbols, Special Characters, Footnotes, or Math in Report Title or Abstract*. (*Abstract*)

Keywords—component, formatting, style, styling, insert (key words)

# Introduction (*Heading 1*)

'Churn' refers to the phenomenon where business clients tend to terminate their services completely or switch to the competitors due to a range of reasons, such as dissatisfaction with service, price hikes, generally appealing offers from competitors or unpredictable circumstances, etc. Each customer defection carries a chain of events with itself, which carries significance for impacting the business's future revenue streams. Customers are the primary assets for any business, losing them means losing your business. Several pieces of research concluded that holding onto your current customer base is far more cost-effective than acquiring new customers. According to estimates, acquiring a new customer could cost 5 to 25 times more, subject to the business’ operating industry.

Current customers being the core asset of most businesses, it's vital to preserve them, as higher customer retention is directly interlinked with market share, recurring revenue and control. Churn, being the metric representing customers ending their relationship with the organisation, can be measured monthly, quarterly or annually. To calculate churn is simple: count your total number of clients at the start of a certain period and compare it to the end of that period. Various industries treat the churn metric differently to retain their customer base efficiently. The default duration for most industries is usually a year, but it varies for companies operating on thin margins or the ones with higher churn rates, as they might trigger a prompt response strategy to counter it. Such as for subscription-based services, SaaS (Software as a Service), gyms, telecom providers, etc.

To implement the appropriate machine learning model, it's crucial to have a deeper understanding of churn. It has two major categories: intentional and involuntary churn. Intentional churn points to changes in customer preferences, like they found a better deal from competitors, poor coverage, not being happy with current services or bad service, etc. It can be even subdivided into deliberate (consumer found an alternative) or incidental (changes in customer circumstances) churn. On the other hand, involuntary churn refers to the complications beyond the customer limits, such as a certain product or service no longer being available, unpaid bills, technical issues or breaches of terms and conditions, etc.

A diagram of people in a circle

AI-generated content may be incorrect.

Figure 1: Illustration of customer base through machine learning.

The primary goal of highly competitive industry these days is to sustain the bloodline of businesses being customers by prolonging their lifespan and gradually increasing their consumption. With the proliferation of the internet and digital services, companies have access to the goldmine of information that is the data. It's crucial to process, refine and optimise the data priorly to make the most out of the customers' behavioural, demographic, psychographic, analytical or transactional data. It wasn’t possible to calculate accurate churn using previous traditional statistical methods, because they can’t handle such high-dimensional complex datasets, but thanks to advancements, it’s possible now through machine learning.

Mainly, churn prediction counters deliberate churn more, as it accounts for the majority of churners. By comprehending the datasets using machine learning models, now businesses have identified hidden patterns and risky and profitable customers. Considering all these factors, companies would be able to have a stronger understanding of their customer base, which will help eventually in the development of customer retention strategies through better deals & offers.

The key purpose of this project is to develop, explore, and evaluate machine learning models. Then apply a state-of-the-art algorithm that can depict customer churn prediction accuracy as close as possible on the chosen dataset. Machine learning provides flexibility in models such as linear/logistic regression, random forest, XG Boost, gradient boosting, CNN or deep learning, etc. Each model can be evaluated using module evaluation metrics (precision, sensitivity, specificity, recall, cross-validation or F1 score) and can be further enhanced through data cleaning, feature selection and hyperparameter tuning to have the best model for churn analysis. Moving beyond, there’s a critical analysis of the derived model, findings, interpretability and ethical considerations for the sensitive customer data and insights.

# Related work

## Selecting a Template (Heading 2)

This template Prepare Your Report Before Styling

Before you begin to format your report, first write and save the content as a separate text file. Complete all content and organisational editing before formatting. Please note sections

An excellent style manual for science writers is [7].

# PROPOSED METHOD

After the text edit has been completed, the report is ready for the template. Duplicate the template file by using the Save As command, and use the naming convention prescribed in the assessment brief for the name of your report. In this newly created file, highlight all of the contents and import your prepared text file. You are now ready to style your report; use the scroll-down window on the left of the MS Word Formatting toolbar.

# Results

kjb

# conclusion and future directions

Hkvk.jn

##### Acknowledgement *(Heading 5)*

Type your acknowledgement here.

##### References

The template will number citations consecutively within brackets [1]. The sentence punctuation follows the bracket [2]. Refer simply to the reference number, as in [3]—do not use “Ref. [3]” or “reference [3]” except at the beginning of a sentence: “Reference [3] was the first ...”

For reports/articles that are used as part of this study and are published in translation journals, please give the English citation first, followed by the original foreign-language citation [6].

1. G. Eason, B. Noble, and I. N. Sneddon, “On certain integrals of Lipschitz-Hankel type involving products of Bessel functions,” Phil. Trans. Roy. Soc. London, vol. A247, pp. 529–551, April 1955. *(references)*
2. J. Clerk Maxwell, A Treatise on Electricity and Magnetism, 3rd ed., vol. 2. Oxford: Clarendon, 1892, pp.68–73.
3. I. S. Jacobs and C. P. Bean, “Fine particles, thin films and exchange anisotropy,” in Magnetism, vol. III, G. T. Rado and H. Suhl, Eds. New York: Academic, 1963, pp. 271–350.
4. K. Elissa, “Title of report if known,” unpublished.
5. R. Nicole, “Title of report with only first word capitalized,” J. Name Stand. Abbrev., in press.
6. Y. Yorozu, M. Hirano, K. Oka, and Y. Tagawa, “Electron spectroscopy studies on magneto-optical media and plastic substrate interface,” IEEE Transl. J. Magn. Japan, vol. 2, pp. 740–741, August 1987 [Digests 9th Annual Conf. Magnetics Japan, p. 301, 1982].
7. M. Young, The Technical Writer’s Handbook. Mill Valley, CA: University Science, 1989.
8. K. Eves and J. Valasek, “Adaptive control for singularly perturbed systems examples,” Code Ocean, Aug. 2023. [Online]. Available: <https://codeocean.com/capsule/4989235/tree>
9. D. P. Kingma and M. Welling, “Auto-encoding variational Bayes,” 2013, arXiv:1312.6114. [Online]. Available: <https://arxiv.org/abs/1312.6114>
10. S. Liu, “Wi-Fi Energy Detection Testbed (12MTC),” 2023, gitHub repository. [Online]. Available: https://github.com/liustone99/Wi-Fi-Energy-Detection-Testbed-12MTC
11. “Treatment episode data set: discharges (TEDS-D): concatenated, 2006 to 2009.” U.S. Department of Health and Human Services, Substance Abuse and Mental Health Services Administration, Office of Applied Studies, August, 2013, DOI:10.3886/ICPSR30122.v2

**This template contains guidance text for composing and formatting reports. Please ensure that all template text is removed before submission. Failure to remove template text from your submission may result in a higher similarity score and may be considered for academic misconduct.**