INTERNATIONAL JOURNAL OF LATEST TECHNOLOGY IN ENGINEERING, MANAGEMENT & APPLIED SCIENCE (IJLTEMAS)

ISSN 2278-2540 | DOI: 10.51583/IJLTEMAS | Special Issue | Volume XIV, Issue XIII, October 2025

Survival Analysis of Customer Lifetime and Churn Prediction in the Telecom Industry

Akshata Lembhe*, Yogita Lagad, Rupali Kamthe, Abhijeet Swami

Department of Statistics, Dr. D. Y. Patil Arts, Commerce and Science College, Pimpri, Pune-18, Maharashtra, India

DOI: https://doi.org/10.51583/IJLTEMAS.2025.1413SP041

Received: 26 June 2025; Accepted: 30 June 2025; Published: 27 October 2025

Abstract: Customer churn poses a significant concern for the telecom industry, as it directly affects both revenue generation and the efficiency of operations. To better understand and address this issue, the present analysis applies survival analysis methods to study customer tenure and the likelihood of churn. Specifically, the Kaplan-Meier estimator is utilized to estimate the survival function of telecom customers over time, while the Cox Proportional Hazards model is used to assess the influence of various customer attributes on the risk of churn. The study highlights that several customer-related factors play a crucial role in determining the probability of churn. Among these, the type of contract (e.g., month-to-month vs. long-term), mode of payment (e.g., electronic check, credit card), and access to additional services (like internet or tech support) emerged as statistically significant determinants. For instance, customers on short-term contracts or using certain payment methods exhibited higher churn probabilities compared to those with long-term commitments or bundled services.

The findings emphasize the importance for telecom companies to tailor their retention strategies by focusing on at-risk customer segments. By understanding the survival patterns and the variables most strongly associated with early churn, service providers can design targeted interventions—such as loyalty programs, contract incentives, or personalized communication—to extend customer relationships and improve overall Customer Lifetime Value (CLV). Ultimately, this evidence-based approach can support telecom firms in minimizing customer loss and maintaining long-term profitability.

Keywords: Survival Analysis, Customer Churn, Kaplan-Meier Estimator, Cox Proportional Hazards Model, Retention Strategies.

I. Introduction

Due to intense competition, the telecom industry faces challenges in retaining customers. Survival analysis, a statistical and machine learning approach, models the time until events like customer churn occur, offering insights into customer behavior and retention strategies. Demographics, usage patterns, service quality, and pricing influence key metrics, such as customer lifetime.

Churn prediction identifies at-risk customers, while survival analysis provides a timeline for churn likelihood, enabling targeted interventions like personalized offers and loyalty programs. Techniques such as Kaplan-Meier estimators and Cox proportional hazards models help estimate survival probabilities and hazard rates, offering actionable intelligence for reducing churn.

By integrating these methods with tools like Power BI, telecom companies can make data-driven decisions, optimize customer retention, and align services with evolving customer expectations. Survival analysis and churn prediction collectively enhance customer relationships, product strategies, and profitability.

II. Literature Review

Survival analysis has been widely used across various domains, including healthcare, engineering, and business, to model time-to-event data. Its application in the telecom industry has gained traction due to its ability to predict customer churn and understand retention dynamics. Studies reveal that techniques like the Kaplan-Meier estimator are effective in estimating survival probabilities, providing insights into customer longevity, and identifying critical churn periods. The Cox Proportional Hazards model further enriches this analysis by quantifying the relationship between multiple predictors and churn risk.

Research highlights several key factors influencing churn, such as contract type, payment method, and additional services. Long-term contracts and automated payment methods have been shown to significantly improve retention, while customers using month-to-month contracts or electronic checks are more likely to churn. Furthermore, studies emphasize the importance of bundling additional services like online security and device protection to enhance customer engagement and reduce attrition.

Recent advancements incorporate machine learning models to complement survival analysis, improving the accuracy of churn predictions. These approaches leverage historical data to identify at-risk customers, allowing companies to implement targeted retention strategies. Integrating survival analysis with visualization tools like Power BI also facilitates better communication of insights, enabling stakeholders to make informed, data-driven decisions. This study builds on existing literature by applying survival analysis to the telecom sector, identifying significant churn factors, and offering actionable recommendations to enhance customer retention.

III. Methodology

The Telco Customer Churn dataset from Kaggle served as the foundation for this study, comprising 7,043 customer records and

ISSN 2278-2540 | DOI: 10.51583/IJLTEMAS | Special Issue | Volume XIV, Issue XIII, October 2025

21 variables, including demographic attributes, service usage patterns, and churn status. Key variables such as tenure, monthly charges, total charges, contract type, payment method, and churn (Yes/No) were analyzed. To ensure the dataset's readiness for analysis, preprocessing steps were undertaken. Missing values were addressed using mean imputation for numerical variables and mode imputation for categorical variables. Categorical variables were encoded using one-hot encoding, while continuous variables, such as Monthly Charges, were normalized to enhance model performance.

Survival analysis was conducted using two key methods. The Kaplan-Meier estimator, a non-parametric technique, was employed to calculate survival probabilities and examine customer churn trends over time. This approach provided insights into the likelihood of customers remaining subscribed at various time intervals. The Cox Proportional Hazards model, a semi-parametric regression technique, was applied to evaluate the influence of covariates on survival time. This model identified significant predictors of churn, such as contract type, payment methods, and additional services, offering actionable insights for retention strategies. By combining these techniques, the study achieved a comprehensive understanding of customer churn and provided a robust framework for predictive analysis.

Dataset-Telco Customer Churn

Dataset link- https://www.kaggle.com/datasets/blastchar/telco-customer-churn

Objective

1) The primary objective of this report is to identify the key factors that influence customer churn in the telecom industry. Understanding what drives customers to leave their service providers is crucial for developing targeted retention strategies and minimizing churn.
2) This study aims to estimate survival probabilities to predict how likely customers are to stay with the service over time. By assessing customer retention patterns, this analysis will help anticipate churn and inform proactive strategies to extend customer lifecycles.
3) This report seeks to understand customer churn behaviour through segment analysis, which examines how different customer groups (based on demographics, service usage, etc.) behave concerning churn. This helps identify specific groups that may require tailored retention efforts.
4) The objective of this study is to visualize key findings using clear and informative visuals. By presenting the results in an intuitive and accessible format, the report aims to make the insights actionable for decision-makers in the telecom industry.
5) This project aims to calculate the Customer Lifetime Value (CLV) for each customer, providing a quantitative measure of the value a customer brings over their entire relationship with the service provider. This information is essential for optimizing customer retention strategies and improving business profitability.

Exploratory Data Analysis

	Senior Citizen	Tenur	Monthly Charges
count	7043.000000	7043.000000	7043.000000
mean	0.162147	32.371149	64.761692
std	0.368612	24.559481	30.090047
min	0.000000	0.000000	18.250000
25%	0.000000	9.000000	35.500000
50%	0.000000	29.000000	70.350000
75%	0.000000	55.000000	89.850000
max	1.000000	72.000000	118.750000

Interpretations:

a) The mean tenure of customers with the telecom service stands at 32.37 months.
b) Customers' tenures span from 0 to 72 months, with the shortest tenure being 0 months and the longest 72 months.
c) A significant number of customers, with an average monthly charge of $64.76, have moderate to high monthly charges, with many paying between $35.50 and $89.85.
d) The proportion of senior citizens among customers averages 16.2%, indicating that a smaller portion of the customer base is made up of senior citizen

ISSN 2278-2540 | DOI: 10.51583/IJLTEMAS | Special Issue | Volume XIV, Issue XIII, October 2025

Data Visualization

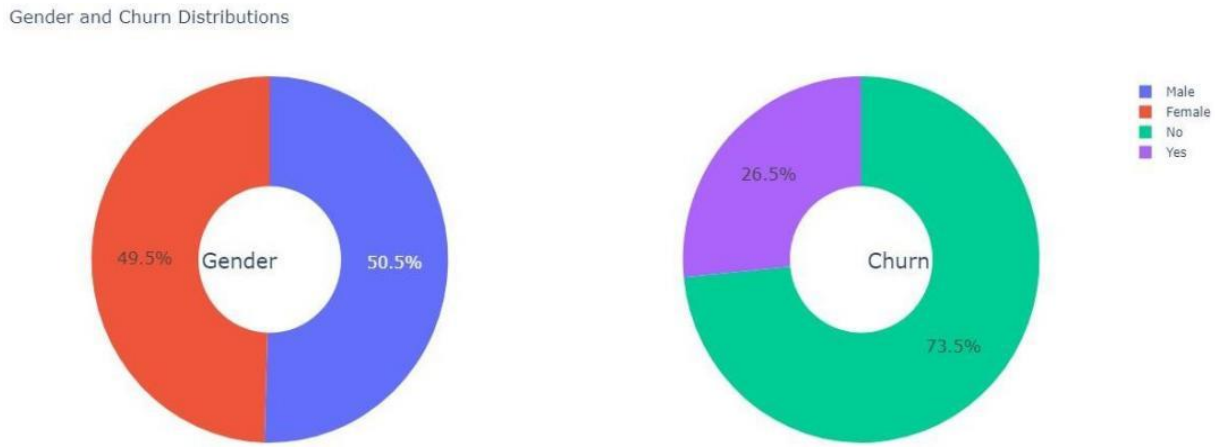


fig1.1 Gender and Churn Distribution

Interpretation: The gender distribution shows no significant bias, which is beneficial for generalizing insights across gender. The churn rate, at 26.5%, suggests that the company might still face challenges in retaining over a quarter of its customers. Strategies to further investigate the reasons behind this churn and mitigate it should be explored.

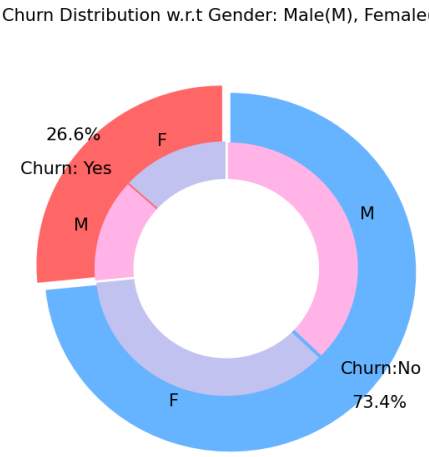


Fig1.2 Gender and Churn Distribution

Interpretation: The analysis reveals that the percentage of customers who switched service providers is nearly equal across genders. Males and females show comparable patterns in their decision to change providers.

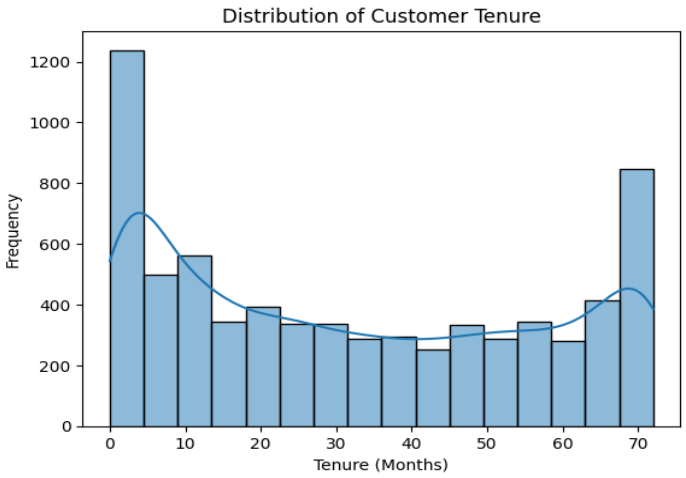


Fig. 2 Tenure Distribution

ISSN 2278-2540 | DOI: 10.51583/IJLTEMAS | Special Issue | Volume XIV, Issue XIII, October 2025

Interpretation: The distribution shows a high number of new customers (short tenure), a decline in mid-tenure retention, and a spike in long-term loyal customers. This suggests retention challenges in the mid-tenure range, with opportunities to improve retention strategies.

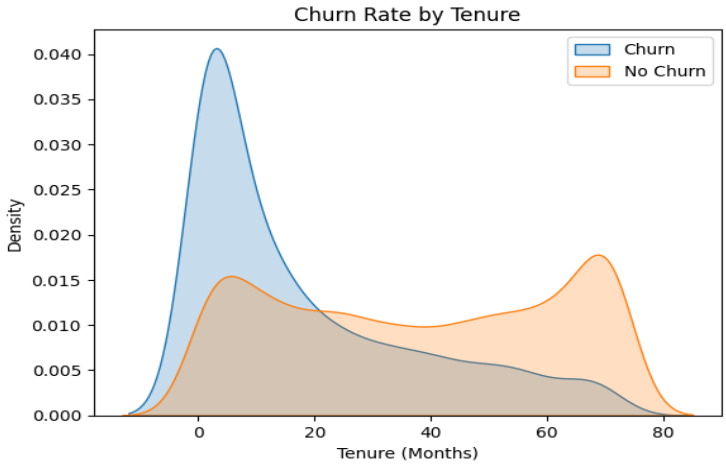


Fig.3 Churn rate by Tenure

Interpretation: Churn is highest within the first 10 months of customer tenure. After this period, the likelihood of customers staying long-term increases, highlighting the importance of early retention efforts.

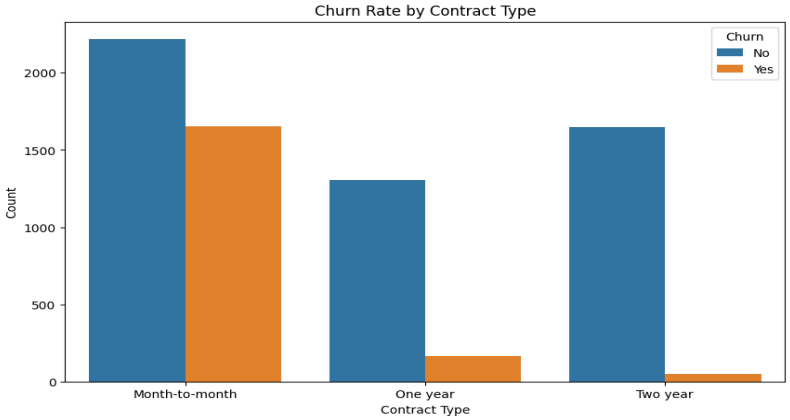
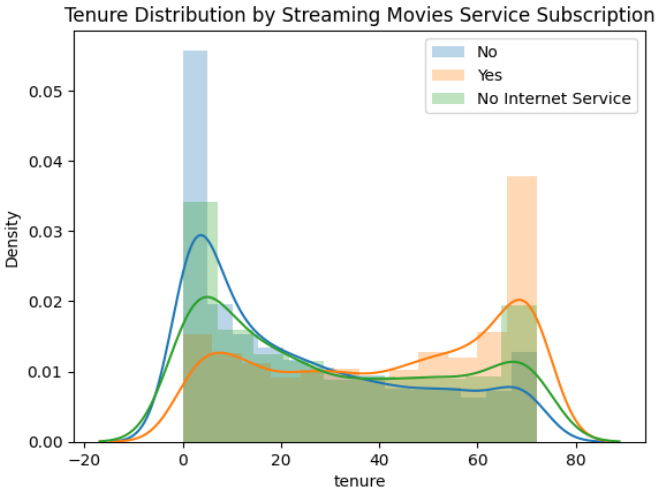


Fig.4 Churn rate by contract type

Interpretation: In the case of Month-to-month contracts Churn rate is very high. There is also a possibility of having customers in the data frame who are still in their two-year or one-year contract plan. About 75% of the customers with a Month-to-Month Contract opted to move out as compared to 13% of customers with One Year Contract and 3% with Two Year Contract.



ISSN 2278-2540 | DOI: 10.51583/IJLTEMAS | Special Issue | Volume XIV, Issue XIII, October 2025

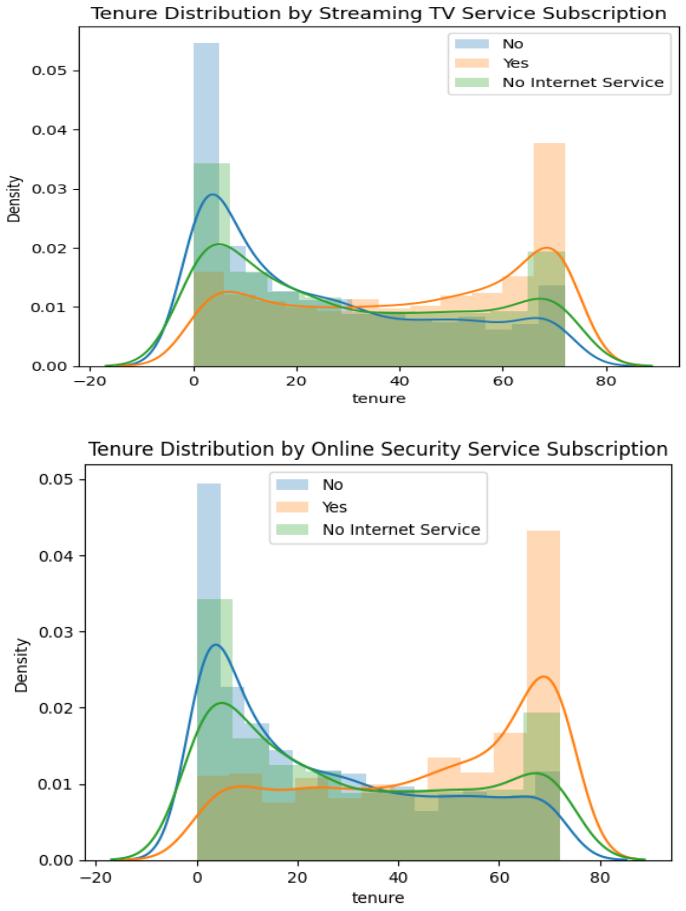


Fig.5 Customer Payment Method distribution w.r.t. Churn

Interpretation: This indicates that new customers who do not subscribe to additional services tend to have higher churn rates. Their limited engagement with the available offerings makes them more likely to leave early in their tenure.

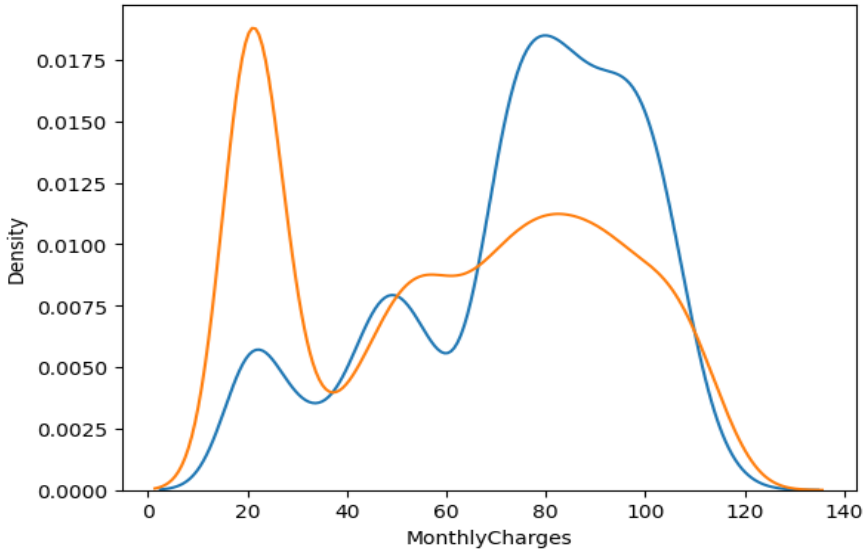


Fig 6.1 Monthly Charges

Interpretation: The density of total charges for churning customers is high around 0. Many customers cancel their subscriptions in 1-2 months.

ISSN 2278-2540 | DOI: 10.51583/IJLTEMAS | Special Issue | Volume XIV, Issue XIII, October 2025

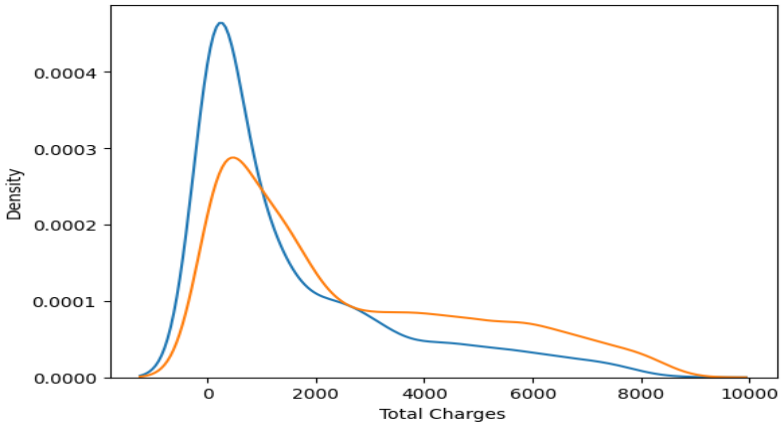


Fig 6.2 Total Charges

Interpretation: The customers paying high monthly fees churn more.

Customer Survival Analysis

Survival Analysis is a statistical approach that studies the time until a particular event of interest takes place. It is widely applied in fields like medicine, engineering, and social sciences to analyse and model the duration until events such as death, system failure, or customer churn occur.

Objectives of Survival Analysis:

To estimate and interpret the survival function.

To compare survival functions across different groups.

To explore the connection between survival time and one or more predictors, we analyse how certain variables influence the duration until an event of interest occurs. This involves assessing the effect of predictors on survival outcomes and identifying key factors that may impact the time frame of the event.

If the event has the probability density function f(t) and cumulative distribution function F(t), then the probability of surviving at least to time t is: Pr(T>t) = S(t) = 1-F(t).

It is a non-increasing function where S(0) = 1 and S(∞) = 0.

Cumulative hazard at time t is defined as H(t) = -ln(S(t)) and instantaneous hazard at time t is h(t)=dH(t)/dt. The instantaneous

hazard can also be written as h(t)=f(t)/S(t)

The likelihood function for survival analysis is described as:

$$l(\beta) = \prod_{i}^{n} [h(ti)^{d}i.\, S(ti)], i = 1, 2.\ldots n$$

where,

di = Censoring variable that equals 1 if the event is observed for individual i and 0 if the event is not observed (censored) for individual i,

h(ti) = Hazard for individual i at time t,

H(ti) = Cumulative hazard for individual i at time t, and

S(ti)= Survival probability for individual i at time t.

Note that when di=0, the contribution of the i'th individual to the likelihood function is just its survival probability until time t: S(t). If the individual has the event, the contribution to the likelihood function is given by the density function f(t)=h(t)S(t).

The log of likelihood is:

$$logl(\beta) = \sum_{i}^{n} [dilog(h(ti)) - H(ti), i = 1, 2.\ldots\ldots n$$

Where log is the natural logarithm.

ISSN 2278-2540 | DOI: 10.51583/IJLTEMAS | Special Issue | Volume XIV, Issue XIII, October 2025

Kaplan-Meier curve

A Kaplan-Meier curve is a graphical representation used to estimate survival probabilities over time for a specific group, frequently employed in medical research. This method is a non-parametric statistic that calculates the survival function based on observed lifetime data.

Advantages and Limitations

The Kaplan-Meier estimator is widely recognized and extensively applied in survival analysis. It is particularly effective for analysing recovery probabilities, mortality rates, and treatment outcomes. However, it has limitations, especially in adjusting survival estimates for covariates. In such cases, parametric survival models or the Cox proportional hazards model can provide more robust analyses by incorporating covariate adjustments.

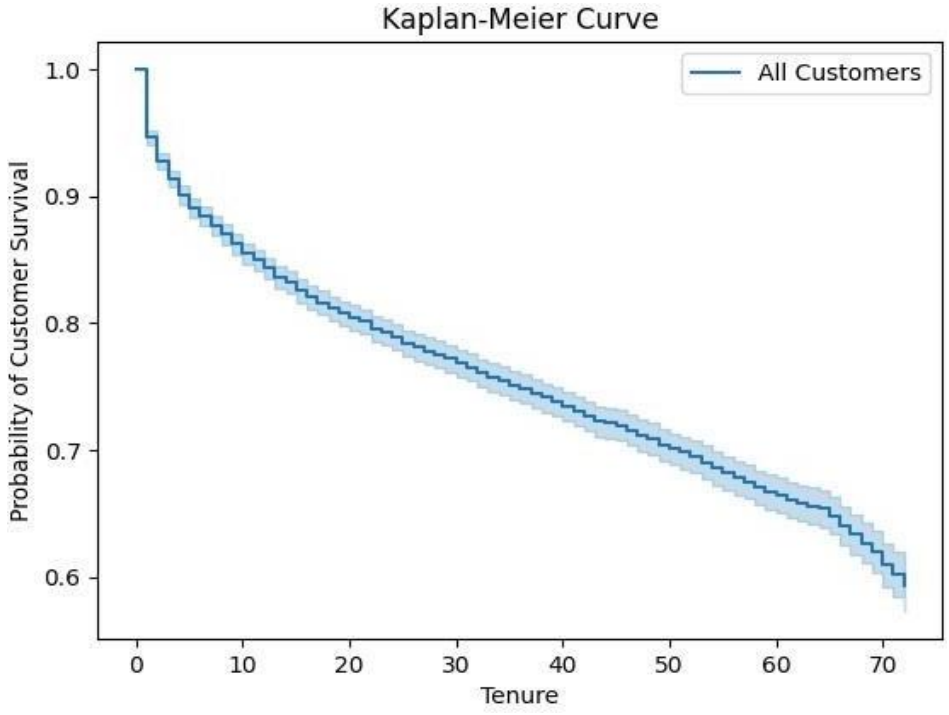


Fig 7. Kaplan-Meier curve

Interpretation: At the beginning, there is a noticeable sharp decline, indicating that customers tend to churn rapidly after completing just one tenure. However, the rate of churn slows down significantly afterwards. As anticipated in the telecom industry, the overall churn rate remains relatively low. The company has successfully retained over 60% of its customers even after 72months.To address this issue, offering greater discounts on long-term plans could be a viable strategy to encourage more customers to opt for extended subscriptions, thereby reducing the initial churn rate.

Log Rank Test

The log-rank test is a statistical method used to compare the survival distributions of two groups. It is a nonparametric hypothesis test, making it suitable for analysing data that is right-skewed and subject to censoring. Importantly, this test assumes that the censoring is non-informative, meaning it is unrelated to the event of interest.

1.Observed Events: At each event time, count the number of events (deaths, failures) in each group.

2.Expected Events: Calculate the expected number of events in each group, assuming that the survival experiences are the same across groups.

The log-rank test statistic is based on the difference between the observed and expected number of events. For two groups, the test statistic can be written as:

$$O = \sum_{i}^{k} \frac{(O1 - E1)^2}{V}$$

Where:

O1: The observed number of events in the group

E1: The expected number of events in group 1.

ISSN 2278-2540 | DOI: 10.51583/IJLTEMAS | Special Issue | Volume XIV, Issue XIII, October 2025

V: The variance of the difference between observed and expected events. K: The number of time points where events occur.

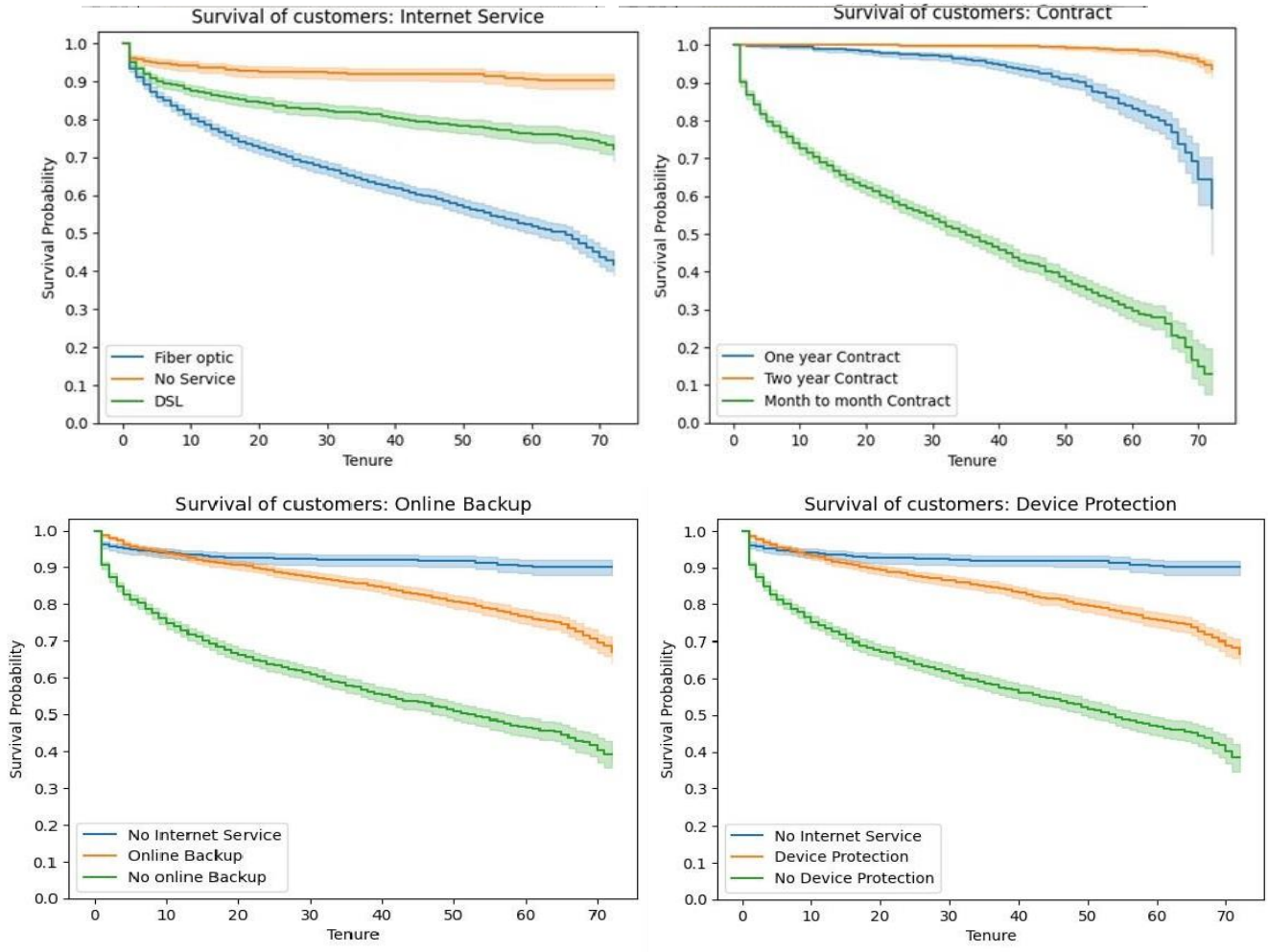


Fig.8 Log Rank Test

Interpretation:

1. Gender: Negligible difference in survival probabilities; gender is not a significant factor in churn.
2. Payment Method: Electronic Check has the highest churn risk, while automated payment methods show better retention.
3. Contract: Month-to-month contracts have the highest churn risk. Longer contracts (One-Year, Two-Year) significantly improve retention.
4. Internet Service: Fiber Optic customers have the highest churn risk compared to DSL and No Service customers.
5. Online Backup: Customers with Online Backup have better retention, while those without face higher churn risk. Promoting online backup can improve customer retention.
6. Device Protection: Customers with Device Protection have higher retention, while those without are more likely to churn. Offering device protection can reduce churn.

C) Survival Regression

Survival Regression extends survival analysis by modelling the relationship between survival time (or time until an event) and explanatory variables (features). For analyzing customer data, I utilize the Cox proportional hazards model to conduct survival regression. This model effectively fits the data, with the resulting coefficients displayed below.

Cox Proportional Hazards Model

This model is a popular choice for examining the relationship between multiple covariates and survival time. It operates under the assumption of proportional hazards, which must be validated as part of the analysis process.

ISSN 2278-2540 | DOI: 10.51583/IJLTEMAS | Special Issue | Volume XIV, Issue XIII, October 2025

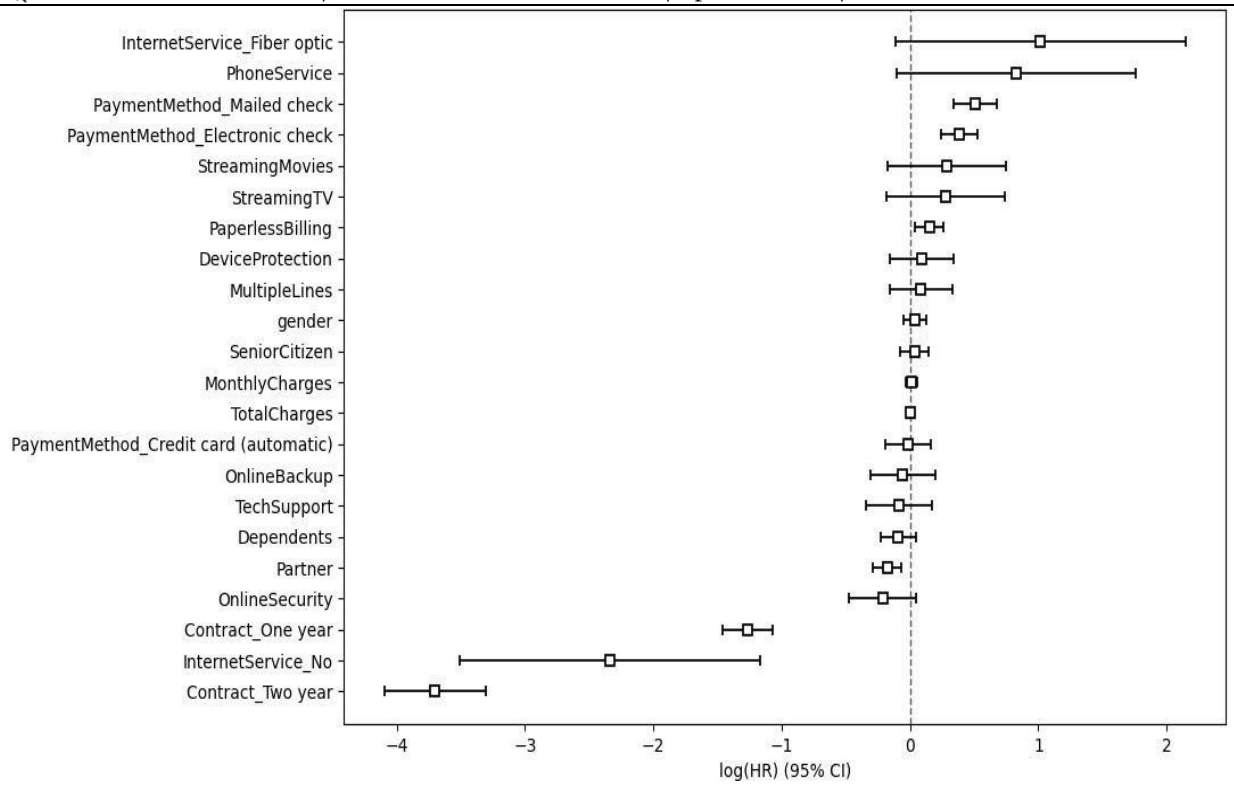


Fig.9 Forest Plot

Forest Plot and Feature Interpretation in the Cox Proportional Hazards Model :

The forest plot displays the log hazard ratios (log(HR)) along with 95% confidence intervals (CI's) for each feature included in the Cox Proportional Hazards Model.

Key Feature Interpretations:

Contract_ Two year and Contract_ One year:

Features with highly negative log(HR) values indicate that one- or two-year contracts significantly lower the risk of churn compared to month-to-month contracts. The confidence intervals are entirely to the left of zero, emphasizing these features as significant predictors with strong protective effects against churn.

Online Security, Tech Support, and Online Backup: These features also have negative log(HR) values, showing that they help reduce the churn risk. Their confidence intervals do not cross zero, further confirming their statistical significance.

Payment Method_ Electronic check:

This feature has a positive log(HR), suggesting that customers using electronic check payments are more likely to churn. Since the confidence interval does not cross zero, it is a significant predictor of failure.

Internet Service_ Fiber optic:

This feature exhibits a highly positive log(HR), indicating that fibre optic internet service greatly increases the churn risk. The hazard ratio is notably large, showing that customers with this service are significantly more prone to churn compared to those using other internet services.

Paperless Billing, Streaming TV, Streaming Movies:

These features have slightly positive log (HR) values, which suggests a marginally higher risk of churn. However, their confidence intervals indicate weaker effects compared to other predictors.

Gender and Senior Citizen:

The confidence intervals for these features are close to or cross zero, implying they are not significant predictors of churn in this model.

Multiple Lines, Device Protection, Phone Service:

These features show log(HR) values near zero or have confidence intervals crossing zero, making them insignificant predictors of

ISSN 2278-2540 | DOI: 10.51583/IJLTEMAS | Special Issue | Volume XIV, Issue XIII, October 2025

churn.

Interpretation:

1)Strong negative predictors (reduce churn):

Long-term contracts (one or two years), online security, tech support, and online backup services are highly effective in reducing churn risk.

2)Strong positive predictors (increase churn):

Fibre optic internet service and electronic check as a payment method are the strongest indicators of higher churn risk.

3)Neutral or insignificant features

like gender, senior citizen status, and certain additional services (e.g., streaming services, multiple lines) have minimal or no significant impact on churn.

C) **Hazard Curve:**

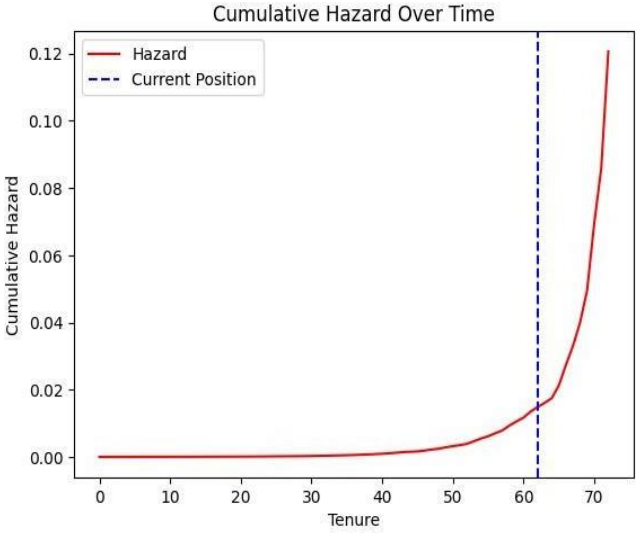


Fig.10 Cumulative Hazard Over Time

Interpretation: A significant increase in hazard rate after 60 months suggests a higher churn likelihood. Targeted

Retention Strategies: Focus on customer engagement and support as customers near the 60-month mark to minimize churn

D) Survival Curve:

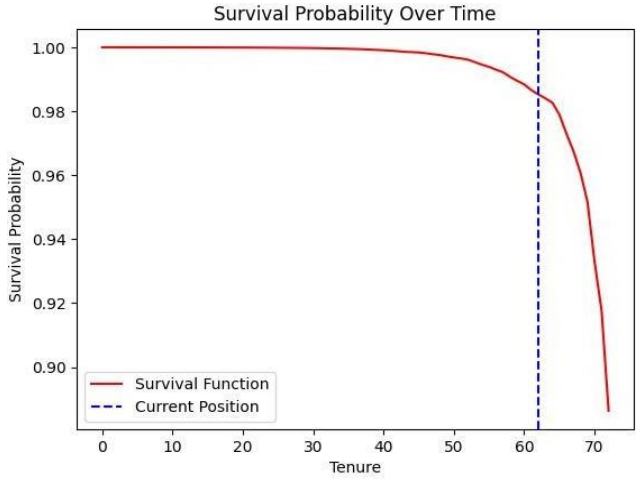


Fig.11 Survival Probability Over Time

Interpretation:

High Early Retention: Survival probability remains near 1 initially. Churn Risk After 60 Months: Significant drop in survival

ISSN 2278-2540 | DOI: 10.51583/IJLTEMAS | Special Issue | Volume XIV, Issue XIII, October 2025

probability, indicating increased churn likelihood.

Conclusion:

Target retention strategies: Focus on customer engagement and satisfaction around the 60-month mark to reduce churn.

Feature Importance

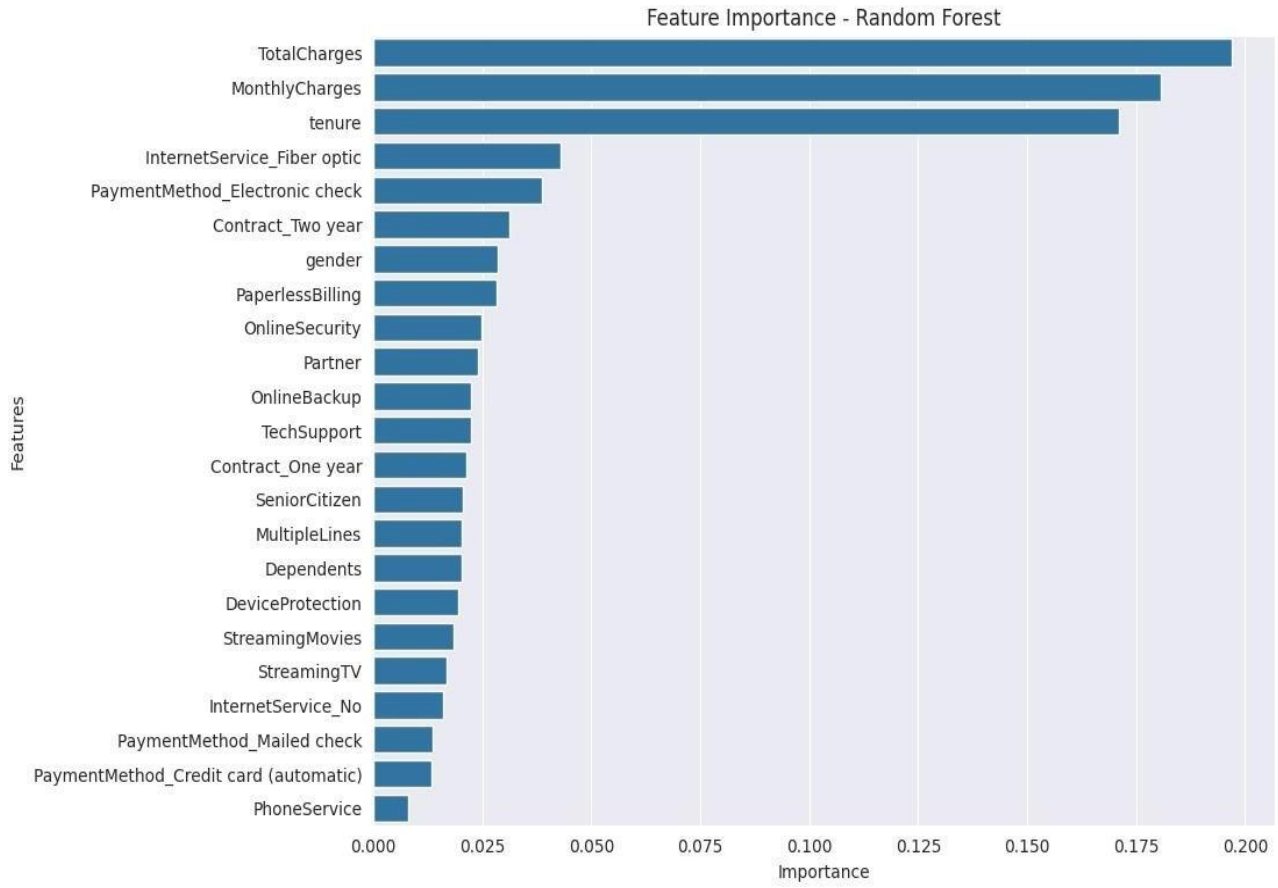


Fig.12 Feature Importance

Interpretation: The graph highlights that monthly charges and how long a customer has been with the company (tenure) are the strongest predictors of customer churn. Customers with higher bills or shorter tenures are more likely to leave. Features like the type of internet service, payment method, and contract length also play a key role in predicting churn, with longer contracts reducing churn risk. Other factors like gender and additional services have less influence.

Customer Lifetime Value

The LTV of a test is: 7344.0 dollars.

LTV Value: The Customer Lifetime Value of $7,344.00 represents the estimated total revenue that a customer is expected to

generate over their lifetime with the telco service.

Interpretation:

High Customer Value: This indicates that each customer in this dataset is potentially very valuable, contributing significantly to the company's overall revenue.

Strategic Investment: Given the substantial LTV, the telco company should focus on customer retention strategies, such as personalized offers, loyalty programs, and enhanced customer service, particularly as customers approach critical churn periods identified in the survival analysis.

Resource Allocation: The high LTV justifies investing resources in improving customer satisfaction and engagement to maximize retention and revenue generation.

Conclusion

The analysis highlights key factors influencing customer churn, including contract type, payment method, and additional services.

ISSN 2278-2540 | DOI: 10.51583/IJLTEMAS | Special Issue | Volume XIV, Issue XIII, October 2025

Survival probability decreases over time, particularly for customers on month-to-month contracts, who are at higher risk of churn. Segment-specific insights reveal that customers using fibre optic internet and electronic check payments are more likely to churn, whereas those with one- or two-year contracts and device protection services have significantly lower churn risk. Additionally, the customer lifetime value has been calculated at $7,344.00, providing a valuable metric for evaluating customer retention strategies.

Retention Focus

To enhance customer retention, the focus should be on several key strategies. Promoting long-term contracts by offering discounts for one- or two-year commitments can encourage customer loyalty. Improving the quality of fibre optic service and introducing value-added services can enhance the overall customer experience. Targeted marketing efforts, including segmented promotions and educational campaigns, can help customers better understand the benefits of staying with the service. Strengthening customer support by providing 24/7 assistance and collecting regular feedback ensures issues are addressed promptly. Implementing loyalty programs with rewards and referral bonuses further incentivizes retention. Proactive churn prevention through predictive analytics can identify at-risk customers, allowing for timely interventions. Fostering community engagement and providing regular updates can help build stronger customer relationships. Offering flexible payment options and incentives for diverse payment methods can also improve satisfaction and retention.

Scope & Limitation

1)Customer Segmentation: Analyze and categorize customers based on demographics, purchasing behaviour, and engagement levels to tailor churn prevention strategies.

2)Predictive Modeling: Develop models to estimate customer lifetime value (CLV) and predict churn risk using historical data.

3)Feature Engineering: Identify key features (e.g., purchase frequency, engagement scores) that impact customer retention and churn likelihood.

4)Data Quality: The analysis depends on the quality of available data, which may include missing values, inaccuracies, or biases that can affect results.

5)Assumptions of Models: Many survival analysis methods rely on assumptions (e.g., proportional hazards) that may not hold in all cases.

6)Temporal Dynamics: Customer behaviour can change over time due to external factors that may not be fully captured in historical dat**a.**

References

1. Kleinbaum, D. G., & Klein, M. (2012). *Survival analysis: A self-learning text* (3rd ed.). Springer.
2. Cleves, M. A., Gould, W. W., Gutierrez, R. G., & Marchenko, Y. V. (2010). *An introduction to survival analysis using Stata* (3rd ed.). Stata Press.
3. Song, L., & Xu, D. (2016). The impact of customer satisfaction on customer loyalty in the telecommunications industry: A comparison of rural and urban areas. *International Journal of Services and Operations Management, 24*(1), 1–22. https://doi.org/10.1504/IJSOM.2016.076285
4. Mishra, A. (2021). Why customer lifetime value matters for your business. *Harvard Business Review*. Retrieved from https://hbr.org
5. McKinsey & Company. (2021). *The future of customer loyalty: A new strategy for a new era*. Retrieved from https://www.mckinsey.com
6. Telco Customer Churn Dataset. (n.d.). *Kaggle*. Retrieved from https://www.kaggle.com/blastchar/telco-customer-churn
7. Hosmer, D. W., Lemeshow, S., & May, S. (2008). *Applied survival analysis: Regression modeling of time-to-event data* (2nd ed.). Wiley-Interscience.
8. Coussement, K., & Van den Poel, D. (2008). Churn prediction in subscription services: An application of support vector machines while comparing two parameter-selection techniques. *Expert Systems with Applications, 34*(1), 313–327. https://doi.org/10.1016/j.eswa.2006.09.038