

Decoding Student Retention and Churn of Vodafone (Telecel) in KNUST: A Survival Analysis Approach

Musah Faridu Oubda (4325620)
Kassim Asana (4323020)
Sarpong Linda (4330720)
Torsi Edmond Collins (4331420)
Asiamah Ezekiel (4316720)



Supervisor By: S. Addai-Henne

Kwame Nkrumah University Of Science And Technology, KNUST
Kumasi, Ghana



TABLE OF CONTENTS

- 1 BACKGROUND OF STUDY
- 2 PROBLEM STATEMENT
- 3 RESEARCH OBJECTIVES
- 4 METHODOLOGY
- 5 ANALYSIS AND FINDINGS
- 6 CONCLUSION
- 7 RECOMMENDATIONS
- 8 REFERENCES



BACKGROUND OF STUDY

- The telecommunication industry is a very competitive market which makes acquiring and retaining new customers challenging.
- Customer churn is the loss of clients or customers [Sterne, 2010]. They include both voluntary and involuntary churn.
- Retention is the practices and strategies a company uses to keep its existing customers.



PROBLEM STATEMENT

- There is a lack of understanding about the factors driving student churn and retention thus making it difficult to develop effective strategies to address this issue [Kapur, 2018].



RESEARCH OBJECTIVES

- To estimate churn rate at various levels.
- To determine covariates that affect the churn rate.
- To identify strategies to improve retention rate and reduce churn.



- **Data collection**

The data used for the study was obtained through simple random sampling of level 400 students in College of Science. The dataset consisted of 338 observations and 14 variables.

- **Data pre-processing**

The data was then encoded to transform all categorical data into a numerical format to facilitate the analysis.



METHODOLOGY CONT'D

Fundamental Concept of Survival Analysis

- **Survival Function $S(t)$**

This is a general framework in survival analysis to determine the probability that a person survives longer than some specified time.

$$S(t) = Pr(T \geq t) \quad (1)$$

- **Hazard Function $h(t)$**

The hazard function denotes the instantaneous rate of failure at time t , given that the subject has survived up to time (t) .

$$h(t) = \lim_{\Delta t \rightarrow 0} \frac{P(t \leq T < t + \Delta t \mid T \geq t)}{\Delta t} \quad (2)$$

$$h(t) \geq 0$$



METHODOLOGY CONT'D

Survival Models

- **Kaplan Meier**

It is to analyze the survival probability of the time until an event occurs.

$$S(t) = \prod_{i:t_i \leq t} \left(1 - \frac{d_i}{n_i}\right) \quad (3)$$

- **Cox Proportional (Cox PH) Hazard Model**

It is used to study the impact of different covariates on survival time.

$$h(t | x) = h_0(t) \exp(\beta_1 x_1 + \beta_2 x_2 + \dots + \beta_p x_p) \quad (4)$$

- **Accelerated Failure Time (AFT)**

It is used to analyze the time until an event occurs by assuming a specific distribution for survival time.

$$\lambda(x) = \exp(b_0 + \sum_{i=1}^n b_i x_i) \quad (5)$$



METHODOLOGY CONT'D

Model Comparison

- **Concordance Index (C-Index)**

It is used to evaluate the predictive accuracy of a survival model.

$$C = \frac{\text{Number of Concordant Pairs}}{\text{Number of Concordant Pairs} + \text{Number of Discordant Pairs}} \quad (6)$$

- **Information Criteria and Loglikelihood**

It is used to evaluate how well a survival model fits the dataset.

AIC	HQIC	BIC
$2k - 2 \ln(L)$	$-2 \ln(L) + 2k \ln(\ln(n))$	$k \ln(n) - 2 \ln(L)$

Table: Information Criteria Formulas



ANALYSIS AND FINDINGS

Descriptive Analysis

Out of the 338 students, 79 were recorded to have churned over the years while 259 had not churned.

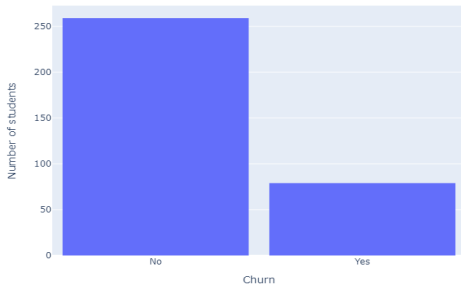


Figure: Descriptive Analysis of Churn Data



ANALYSIS AND FINDINGS CONT'D

Kaplan Meier

The Kaplan-Meier survival analysis below estimates the survival probability at the end of various academic years.

Event Time	Number of Events	Survival Probability
0	0	1.000000
1	25	0.926036
2	22	0.860947
3	32	0.766272

Table: Survival Analysis Summary



ANALYSIS AND FINDINGS CONT'D

Kaplan Meier Curve

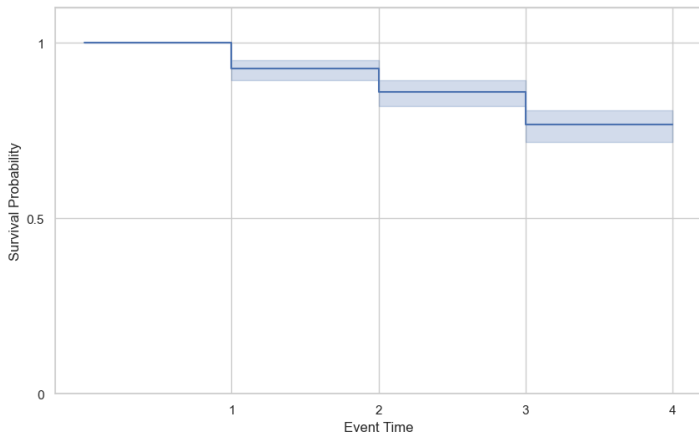


Figure: Kaplan Meier Curve



ANALYSIS AND FINDINGS CONT'D

Accelerated Failure Time

Selection of parametric distribution

The LogNormal model is the best based on AIC and BIC values, showing the best fit among the distributions.

Model	AIC	BIC	Hannan-Quinn
Weibull	316.43	300.08	306.67
LogNormal	305.81	289.45	306.67
LogLogistic	307.51	291.16	306.67

Table: Information Criteria on AFT Models



ANALYSIS AND FINDINGS CONT'D

Comparison between Cox PH and AFT

* indicates significant covariate.

Variable	Cox PH Model		Lognormal Model	
	β	P-values	β	P-values
Gender	0.33	0.24	0.032	0.785
Residence	0.32	0.25	-0.146	0.239
Usage Freq.	-0.08	0.27	-0.002	0.940
Voice Calls	-0.57	0.08	0.132	0.329
Mobile Data	0.32	0.48	-0.287	0.165
SMS	0.48	0.06	-0.228	0.036*
Data Exhaust.	-0.20	0.61	0.102	0.493
Multiple Networks	-0.13	0.78	-0.104	0.607
Poor Network	2.60	<0.005*	-1.041	<0.0005*
Customer Service	-1.40	<0.005*	0.809	<0.0005*
High Costs	-1.13	<0.005*	0.515	<0.0005*
Monthly Data	0.21	0.02*	-0.069	0.113
Intercept μ	–	–	1.626	<0.0005
Intercept σ	–	–	-0.616	<0.0005

Table: Cox PH and LogNormal Models



ANALYSIS AND FINDINGS

Cox Proportional Hazard Plot

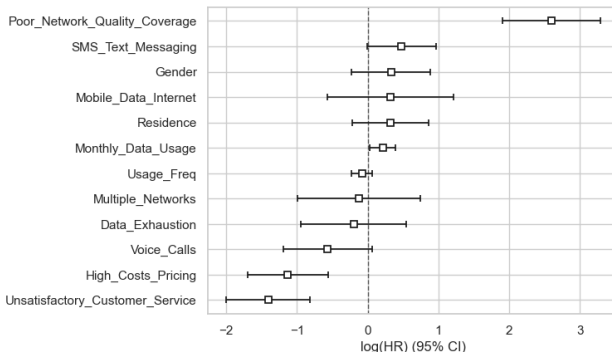


Figure: Cox PH Forest Plot

A positive covariate indicates an increase in churn rate while the negative covariate indicates a decrease in churn rate.



ANALYSIS AND FINDINGS CONT'D

Accelerated Failure Time Plot

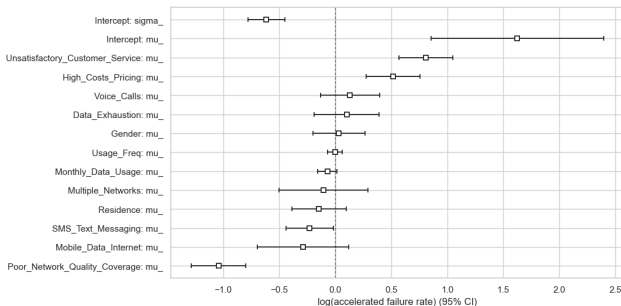


Figure: LogNormal Forest Plot

A positive covariate indicates a decrease in the time to event while the negative covariate indicates an increase in the time to event.



MODEL COMPARISON

Model	Concordance	Loglikelihood
LogNormal	0.958	290.678
Cox PH	0.96	266.68

Table: Model Concordance and Loglikelihood values



CONCLUSION

- The research indicates that the Cox PH model provides the best predictions for the data, while the LogNormal model offers the best fit.
- Poor Network, Customer Service, High Costs, Monthly Data, and SMS are the most significant covariates in the study.
- Poor Network Quality tends to increase both the churn rate and time to churn the most.
- Students tend to churn most in their 3rd year.



RECOMMENDATIONS

- Enhance the quality and reliability of network coverage across KNUST to reduce churn rates.
- Establishment of avenues and platforms to address students' concerns more effectively.



REFERENCES

[Kapur, 2018] Kapur, R. (2018).

Factors driving customer churn in telecom industry.

International Journal of Telecom Studies, 22(4):321–334.

[Sterne, 2010] Sterne, J. (2010).

Social Media Metrics: How to Measure and Optimize Your Marketing Investment.

Wiley.



THANK YOU

