Paper 3501 - 2015

Predicting transformer lifetime using survival analysis and modeling risk associated with overloaded transformers Using SAS® Enterprise Miner™ 12.1

Balamurugan Mohan and Dr. Goutam Chakraborty, Oklahoma State University

ABSTRACT

Utility companies in America are always challenged when it comes to knowing when their infrastructure fails. One of the most critical components of a utility company's infrastructure is the transformer. It is important to assess the remaining lifetime of transformers so that the company can reduce costs, plan expenditures in advance and mitigate the risk of failure to a large extent. It is also equally important to identify the high risk transformers in advance and maintain them accordingly to avoid sudden loss of equipment due to overloading. The objective of this paper is to use SAS® to predict the lifetime of transformers, identify the various factors that contribute to their failure and model the transformer into high, medium and low risky categories based on load for easy maintenance. The data set used in this study is from a utility company in Southwestern United States and contains around 18,000 observations and 26 variables from 2006 till 2013. Survival analysis is performed on this data. By building a Cox's regression model, the important factors contributing to the failure of a transformer are identified. Several risk based models are then built to categorize the transformers into high, medium and low risk categories based on their loads.

INTRODUCTION

Power transformers are important assets in the utility industry. The loss due to the failure of a transformer comprises of costs to replace the transformer as well as the bad reputation of reliability and lack of service to customers. In commercial environments, the transformers are operated in high power environment, hence chances of their failing due to severe overloading may be higher. The average age of a power transformer is 40 years. The Utility Company began storing the age of the transformers from 2006. Hence, our survival analysis is to predict life time for these transformers that were installed after 2006.

The main areas of research in this paper are as follows:

- Use survival analysis to predict the life time of transformers
- Build non parametric models for failure time data to explore lifetime of transformers based on age and overloaded strata
- Find important factors that contribute to the failure of the transformer using Cox's Proportional hazard model
- Build transformer risk based models to identify the transformers that get overloaded in advance so that these can be maintained properly

RIGHT CENSORED DATA

Right censoring occurs when a subject leaves the study before an event occurs, or the study ends before the event has occurred. In this study we are looking at a time period of 2006 - 2013. Hence, the transformers that have not failed after 2013 are considered to be right censored. In our analysis, we have considered data from 2006 to 2013 as shown in the figure below.

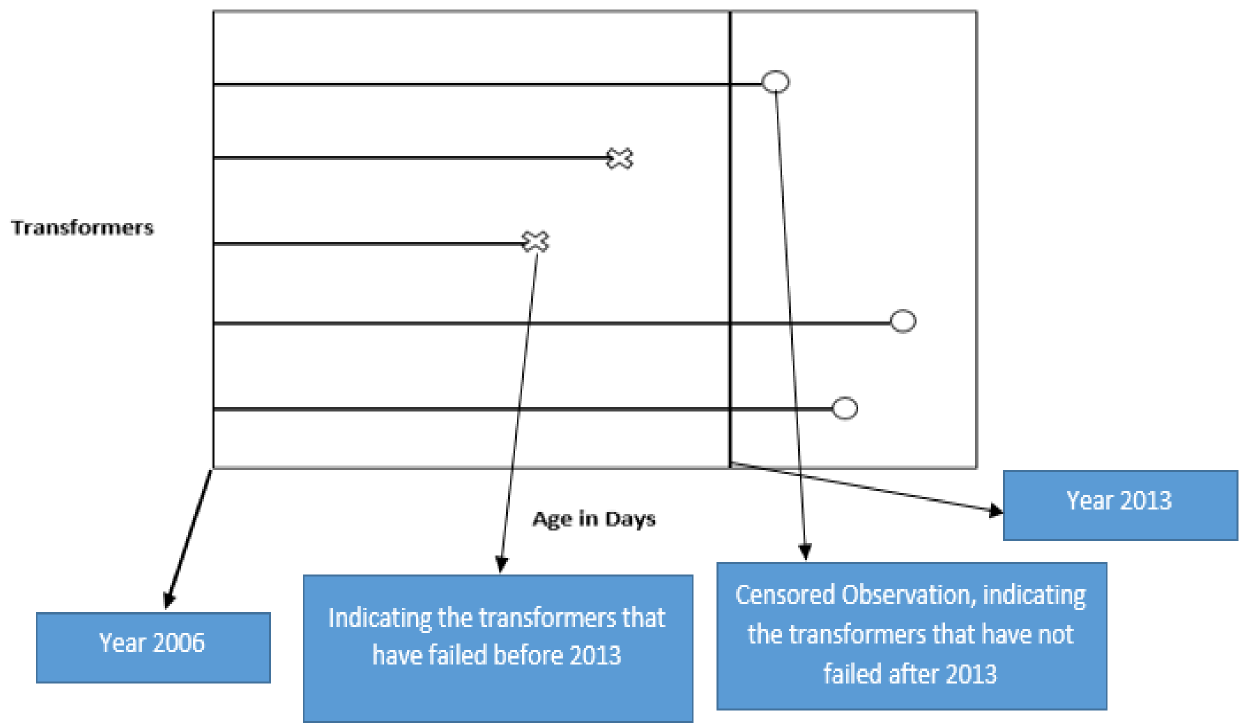

Figure 1. Right censored data

DATA COLLECTION AND ANALYSIS

The data was obtained by merging 12 different tables (as shown in Figure 2). The current table (transformer details), the load table (7 different tables containing information on load, temperature, kVA rating and other factors from 2006 till 2013), the failures table (failure information of transformers), normal and overloaded transformer conditions table were merged together to form one final flat file for the analysis data set.

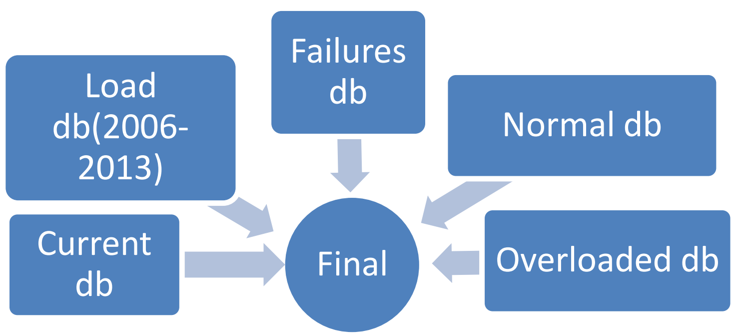

Figure 2. Data consolidation schematic view

Num.	Variable	Type of Variable	Description
1	Age	Continuous/ Response	Age of the transformer in days
2	Avg_temp_f	Continuous	Average maximum temperature of the transformer
3	Avg_Loaded_Max	Continuous	Average maximum load of the transformer
4	Avg_kVA_Rating	Continuous	Average kVA rating of the transformer
5	Normal	Continuous	Number of times the transformer was normal
6	Overloaded	Continuous	Number of times the transformer was overloaded
7	Indicator	Binary / Censor	1 – Old (Failed) 0 – New (Existing)
8	Cat_Ind	Binary	1 - Commercial 0 - Residential

Table 1. Final variables in the analysis data set

The final data set (Table 1) has variables: age, average kVA, average temperature, average load, Indicator=1 (Failed Transformers), Indicator=0 (Censored variables – existing transformers), normal and overloaded Conditions for residential (Cat_Ind=0) and commercial transformers (Cat_Ind=1). From the bar chart, we can see that there are more number of residential transformers compared to the commercial transformers (Figure 3) in the data.

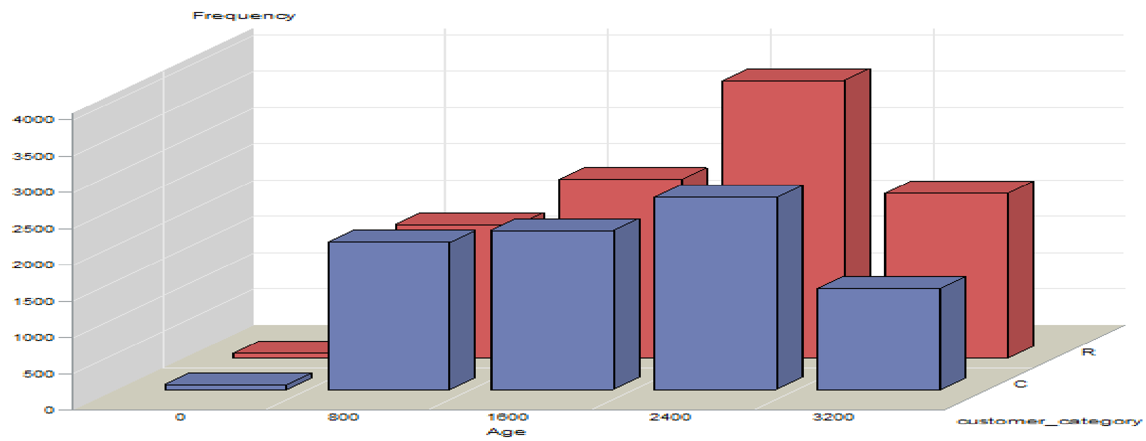

Figure 3. Age of the transformer based on category

PROBABILITY DENSITY FUNCTION OF FAILED TRANSFORMERS

Consider a random variable, time, which records survival times. The function that describes likelihood of observing time at time t relative to all other survival times is known as the probability density function (pdf), or f (t). Integrating the pdf over a range of survival times gives the probability of observing a survival time within that interval.

$$\Pr[a \le X \le b] = \int_a^b f_X(x)\,dx.$$

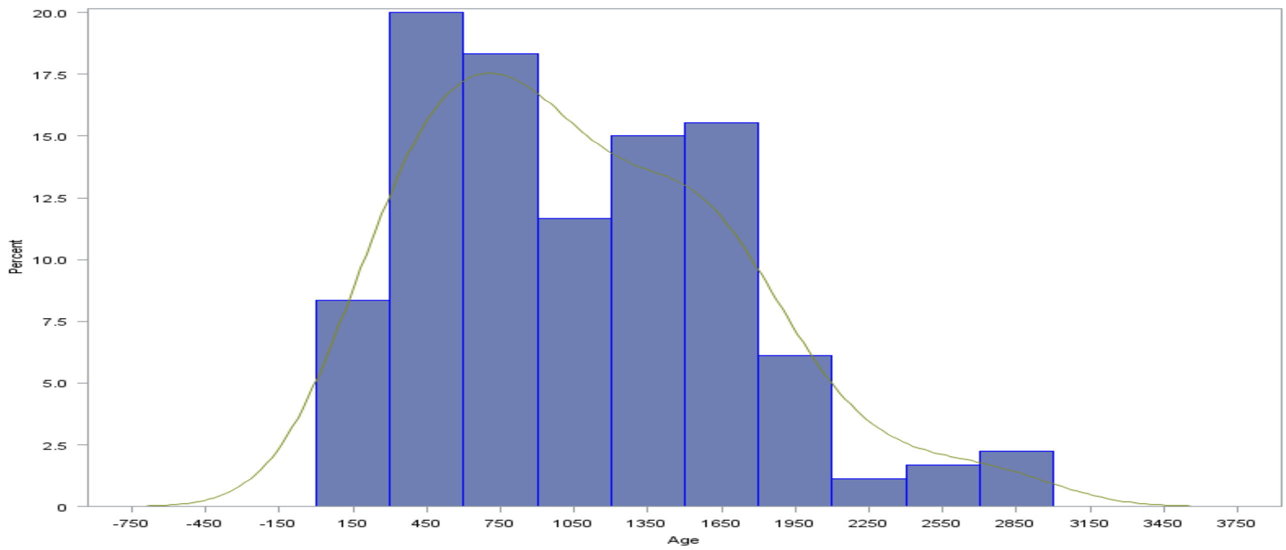

Figure 4. Probability density function of failed transformers

We can see that risk of transformer failure is higher between the ages of 450-1,650 days and then this value decreases drastically as shown in the figure above.

CUMULATIVE DISTRIBUTIVE FUNCTION OF FAILED TRANSFORMERS

The cumulative distribution function (cdf), F (t), describes the probability of observing time less than or equal to sometime t, or P (Time≤t). The cumulative distribution function is shown below:

$$F(t) = \int_{0}^{t} f(t)\, dt$$

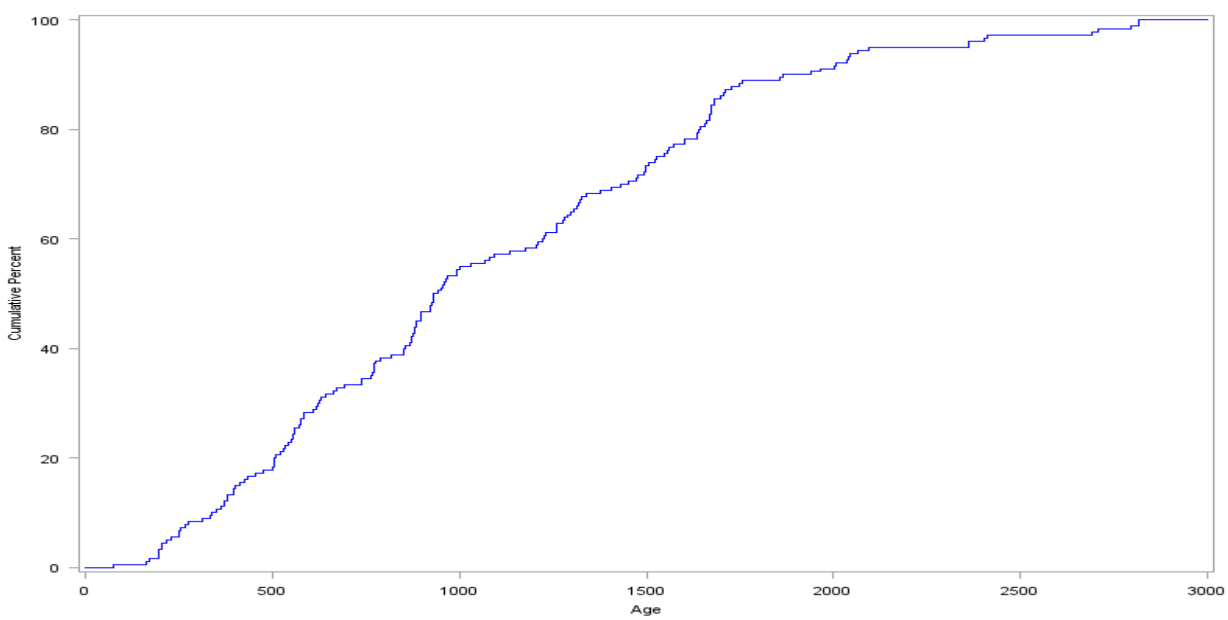

Figure 5. Cumulative distributive function of failed transformers

We can see that probability of transformer surviving till 1,500 days is higher than 50% as shown above.

SURVIVAL FUNCTION OF FAILED TRANSFORMERS

A simple transformation of the cumulative distribution function produces the survival function, S (t):

$$S\ (t) = 1 - F\ (T)$$

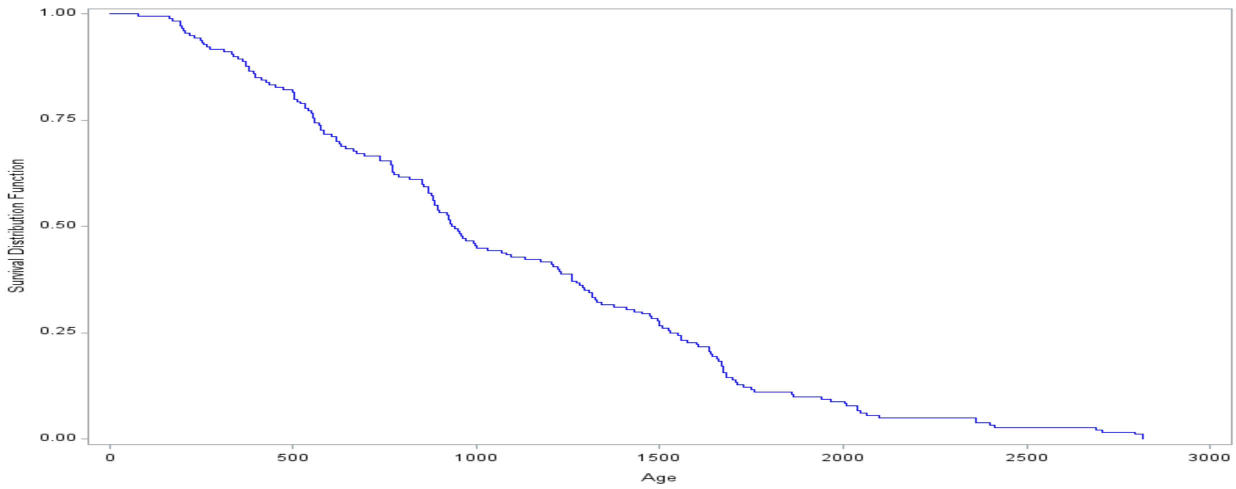

Figure 6. Survival function of failed transformers

We can see that the probability of a transformer surviving more than 2,000 days is about 10% as shown above.

CORRELATION

Pearson Correlation Coefficients, N = 18785 Prob > |r| under H0: Rho=0				
	Avg_kva_Rating	Age	Avg_Temp_f	Avg_Loaded_Max
Avg_kva_Rating	1.00000	-0.03152	0.06080	-0.03936
		<.0001	<.0001	<.0001
Age	-0.03152	1.00000	0.16995	0.10178
	<.0001		<.0001	<.0001
Avg_Temp_f	0.06080	0.16995	1.00000	0.07091
	<.0001	<.0001		<.0001
Avg_Loaded_Max	-0.03936	0.10178	0.07091	1.00000
	<.0001	<.0001	<.0001	

Table 2. Correlation analysis

Pearson correlation analysis is performed using PROC CORR and results show that there is not to high multicollinearity among the interval input variables.

KAPLAN-MEIER SURVIVAL ESTIMATION

This method (also known as product-limit method) produces an estimate of the survival function based on complete or censored data.

$$\hat{S}(t) = \prod_{t_i \le t} \frac{n_i - d_i}{n_i} ,$$

Where, n_i is the number of subjects at risk and d_i is the number of subjects who fail, both at time t_i. Thus, each term in the product is the conditional probability of survival beyond time t_i, meaning the probability of surviving beyond time t_i, given the subject has survived up to time t_i. The survival function estimate of the unconditional probability of survival beyond time t (the probability of survival beyond time t from the onset of risk) is then obtained by multiplying together these conditional probabilities up to time t together.

The LIFETEST Procedure

Product-Limit Survival Estimates

Age		Survival	Failure	Survival Standard Error	Number Failed	Number Left
0.00		1.0000	0	0	0	18785
66.00	*	.	.	.	0	18784
75.00		0.9999	0.000053	0.000053	1	18783
160.00	*	.	.	.	1	18782
162.00		0.9999	0.000106	0.000075	2	18781
168.00	*	.	.	.	2	18780
171.00		0.9998	0.000160	0.000092	3	18779
177.00	*	.	.	.	3	18778
182.00	*	.	.	.	3	18777
185.00	*	.	.	.	3	18776
193.00		.	.	.	4	18775
193.00		0.9997	0.000266	0.000119	5	18774
194.00		0.9997	0.000319	0.000130	6	18773
195.00	*	.	.	.	6	18772
199.00	*	.	.	.	6	18771
201.00	*	.	.	.	6	18770
202.00		0.9996	0.000373	0.000141	7	18769
205.00		0.9996	0.000426	0.000151	8	18768
211.00	*	.	.	.	8	18767
212.00	*	.	.	.	8	18766
216.00		0.9995	0.000479	0.000160	9	18765

Table 3. Product-limit survival estimates

From the above Product-Limit Estimates we find that there are few failures between 0 - 200 days.

APPLYING LIFE-TABLE METHOD

Life Table Method allows to estimate survival, probability density and failure rate functions from complete or censored data.

The LIFETEST Procedure

Life Table Survival Estimates

Interval												Evaluated at the Midpoint of the Interval			
[Lower,	Upper)	Number Failed	Number Censored	Effective Sample Size	Conditional Probability of Failure	Conditional Probability Standard Error	Survival	Failure	Survival Standard Error	Median Residual Lifetime	Median Standard Error	PDF	PDF Standard Error	Hazard	Hazard Standard Error
0	500	33	341	18614.5	0.00177	0.000308	1.0000	0	0	.	.	3.546E-6	6.167E-7	3.549E-6	6.178E-7
500	1000	65	2599	17111.5	0.00380	0.000470	0.9982	0.00177	0.000308	.	.	7.584E-6	9.389E-7	7.612E-6	9.441E-7
1000	1500	34	2436	14529.0	0.00234	0.000401	0.9944	0.00556	0.000561	.	.	4.654E-6	7.973E-7	4.686E-6	8.036E-7
1500	2000	32	3109	11722.5	0.00273	0.000482	0.9921	0.00789	0.000687	.	.	5.417E-6	9.562E-7	5.467E-6	9.664E-7
2000	2500	11	4153	8059.5	0.00136	0.000411	0.9894	0.0106	0.000836	.	.	2.701E-6	8.138E-7	2.732E-6	8.236E-7
2500	3000	5	4501	3721.5	0.00134	0.000600	0.9880	0.0120	0.000928	.	.	2.655E-6	1.187E-6	2.689E-6	1.203E-6
3000	3500	0	1466	733.0	0	0	0.9867	0.0133	0.00110	.	.	0	.	0	.
3500	.	0	0	0.0	0	0	0.9867	0.0133	0.00110	.	.	.	.	.	.

Table 4. Life-table survival estimates

From, the above life table survival estimates, we find that there is more number of failures in the age interval between 500 – 1,000 days and in the interval between 1,500 – 2,000 days. There are fewer number of failures after the age of 2,000 days.

PROC LIFE TEST BASED ANALYSIS

AGE BASED STRATUM

PROC LIFETEST can be used to compute the product-limit estimate of the survivor function for each treatment and to compare the survivor functions between the two treatments. PROC LIFETEST is invoked here to compute the product-limit estimate of the survivor function for each transformer category and to compare the survivor functions between the two categories.

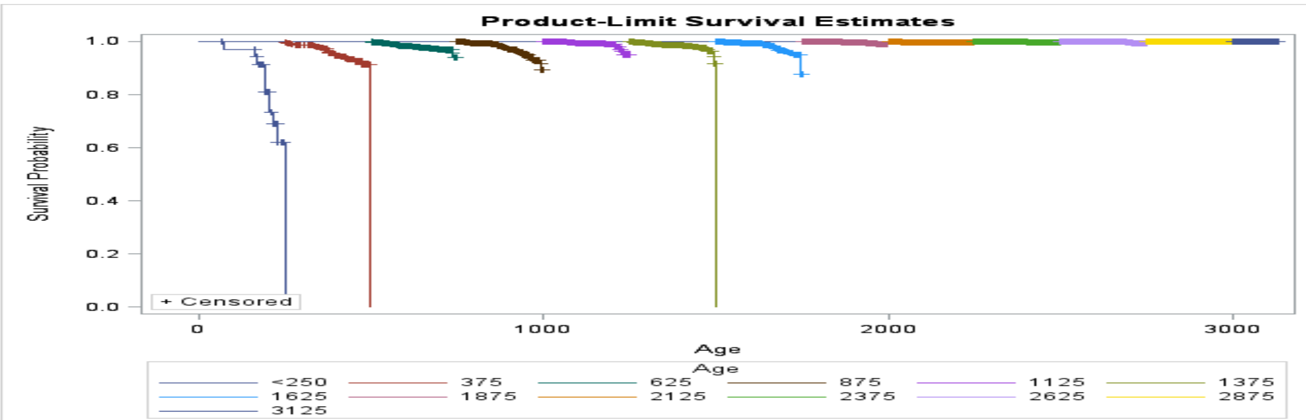

Figure 7. Product-limit survivor plot

Using PROC LIFETEST across different (Figure. 7) age based strata (Intervals of 250 days from 0 – 3,500 days), we are able to see a greater decline in survival probabilities from the age of 0 to 1,500 days. This gives us an interesting insight that if a transformer is going to fail in a short period of time, it is going to predominantly fail in the first 1,500 days.

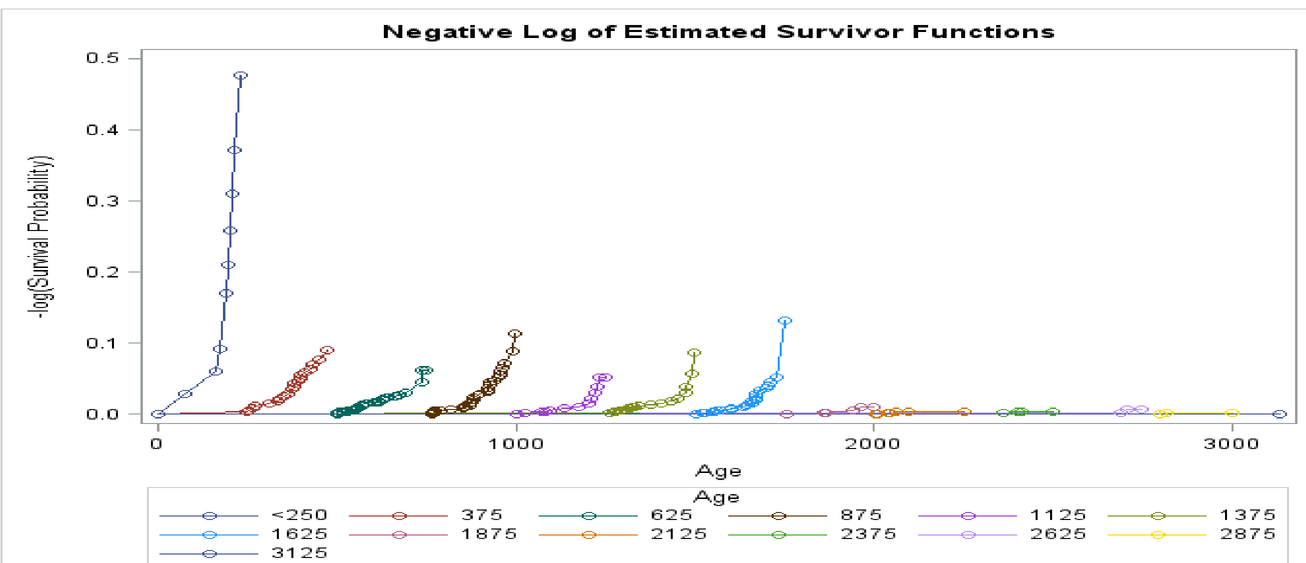

Figure 8. Negative Log survival plot

The Negative log survival plot (Figure. 8) shows us that from 0 to 1,625 days, the survival function seems to be straight. This in turn suggest to us that in this stratum the hazard function also increases comparing to the other strata. This re-iterates our previous result on short lived transformers.

OVERLOAD BASED STRATUM

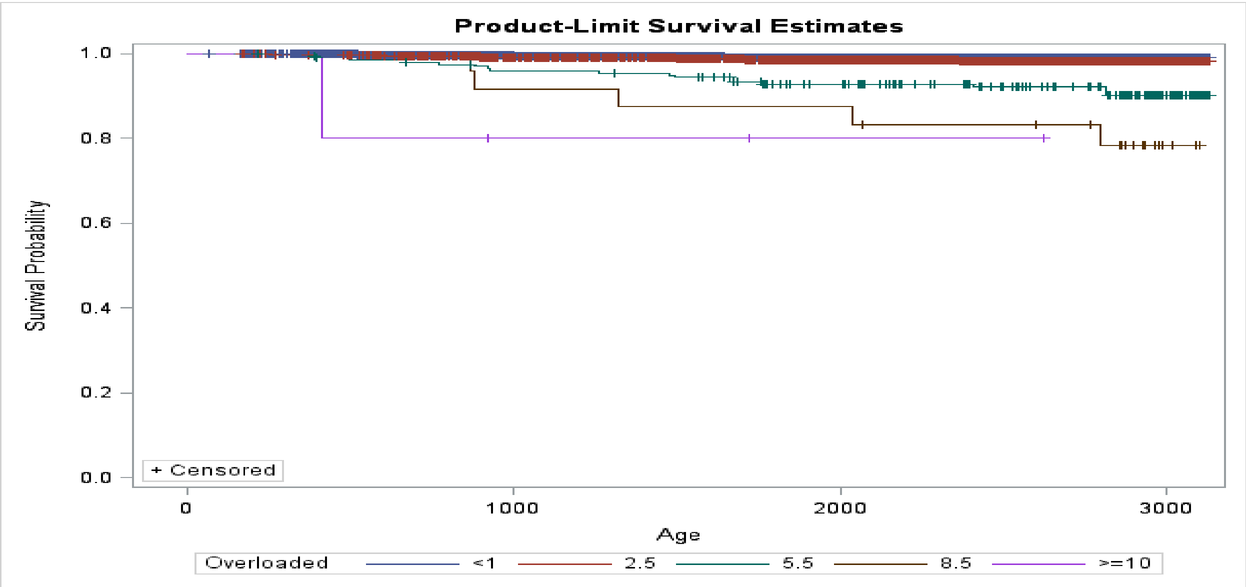

Figure 9. Product-limit survivor functions

Using PROC LIFETEST across different (Figure 9) overload based strata (Intervals of 2 from 0 to 20 times), we are able to see that the survival probabilities decreases as the number of times a transformer gets overloaded increases.

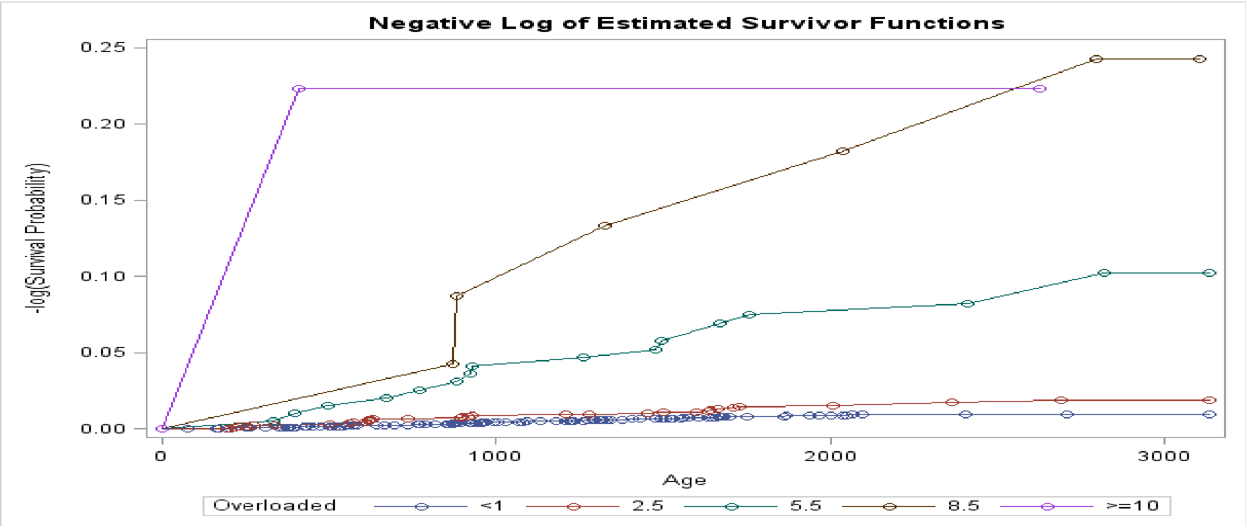

Figure 10. Log – Log survival functions

The Log negative log survival plot (Figure. 10) shows us that the survival probability of commercial transformers decreases faster when compared to those of the residential transformers.

SMOOTH HAZARD FUNCTION

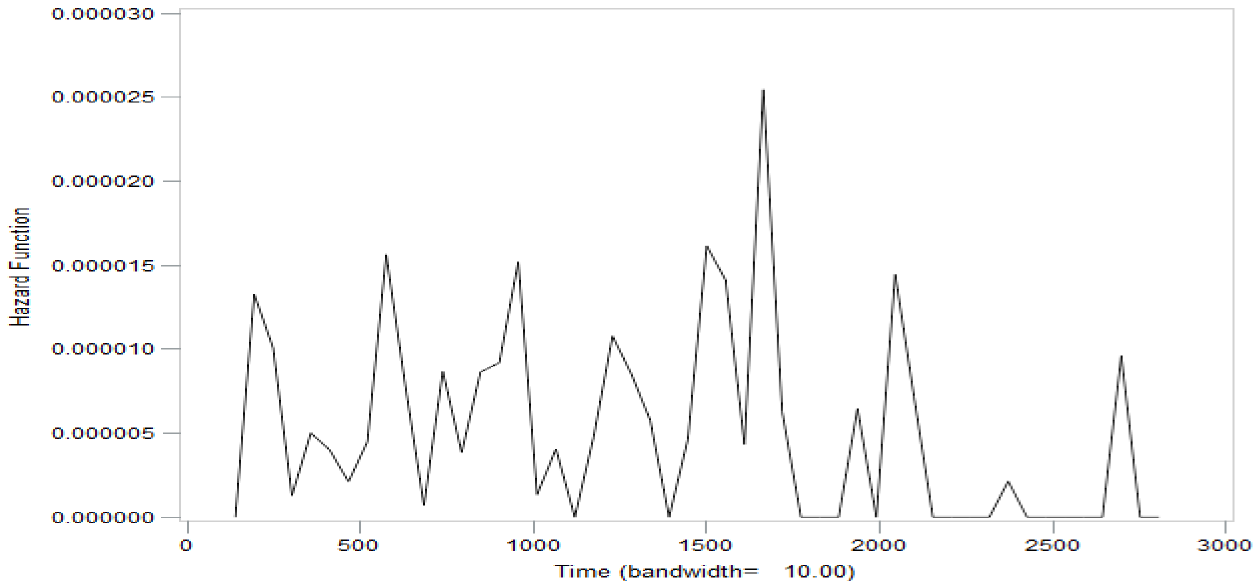

Figure 11. Smoothed hazard function

We used a SMOOTH macro that produces non parametric plots of hazard functions using a kernel smoothing method (Figure. 11). The macro uses the data set from the OUTSRV statement of the PROC LIFETEST. From the output hazard function we are able to see that there are considerable amount of peaks at the time period of 1,500-1,625 days.

PROC PHREG BASED ANALYSIS – SCORE BASED MODEL

The PHREG Procedure

Regression Models Selected by Score Criterion		
Number of Variables	Score Chi-Square	Variables Included in Model
1	226.7617	Overloaded
1	42.5868	Normal
1	41.7839	Avg_Loaded_Max
2	249.8610	Avg_Temp_f Overloaded
2	249.6223	Normal Overloaded
2	229.6534	Avg_kva_Rating Overloaded
3	274.7438	Avg_Temp_f Normal Overloaded
3	253.7829	Avg_kva_Rating Avg_Temp_f Overloaded
3	253.2306	Avg_kva_Rating Normal Overloaded
4	279.5888	Avg_kva_Rating Avg_Temp_f Normal Overloaded
4	275.1475	Avg_Temp_f Avg_Loaded_Max Normal Overloaded
4	254.3576	Avg_kva_Rating Avg_Temp_f Avg_Loaded_Max Overloaded
5	280.1179	Avg_kva_Rating Avg_Temp_f Avg_Loaded_Max Normal Overloaded

Table 5. Regression models selected by score criterion

The Variable importance is judged through Chi-Square based score criterion (PHREG model).

COX'S PROPORTIONAL HAZARD MODEL

In the Cox's proportional hazard model, the response variable age is crossed with the censoring variable (Indicator = 0 for existing transformers). The other variables used in the model are Overloaded, Normal, Avg_temp_f, Avg _kVA_Rating and Avg_Loaded_Max.

Summary of the Number of Event and Censored Values			
Total	Event	Censored	Percent Censored
18785	180	18605	99.04

Convergence Status
Convergence criterion (GCONV=1E-8) satisfied.

Model Fit Statistics		
Criterion	Without Covariates	With Covariates
-2 LOG L	3425.523	3274.065
AIC	3425.523	3286.065
SBC	3425.523	3305.223

Testing Global Null Hypothesis: BETA=0			
Test	Chi-Square	DF	Pr > ChiSq
Likelihood Ratio	151.4577	6	<.0001
Score	280.1528	6	<.0001
Wald	221.3623	6	<.0001

Analysis of Maximum Likelihood Estimates						
Parameter	DF	Parameter Estimate	Standard Error	Chi-Square	Pr > ChiSq	Hazard Ratio
Avg_kva_Rating	1	-0.00178	0.0008748	4.1341	0.0420	0.998
Avg_Temp_f	1	0.03700	0.00706	27.4942	<.0001	1.038
Avg_Loaded_Max	1	0.0003379	0.0005746	0.3457	0.5565	1.000
Cat_Ind	1	-0.00657	0.15845	0.0017	0.9669	0.993
Normal	1	-0.20182	0.03984	25.6564	<.0001	0.817
Overloaded	1	0.28811	0.03301	76.1831	<.0001	1.334

Table 6. PHREG model results

From the PHREG model results with TIES=DISCRETE option we can see that Overloaded, Normal, Avg_temp_f and Avg _kVA_Rating (with $p<0.05$) are the major factors contributing to the hazard of the transformers (Table 6). From the SMOOTH hazard macro (Figure 12) that produces non parametric plots of hazard functions using a kernel smoothing method, we can see that Group 1 (commercial transformers) has a higher hazard rate when compared to Group 0 (residential transformers).

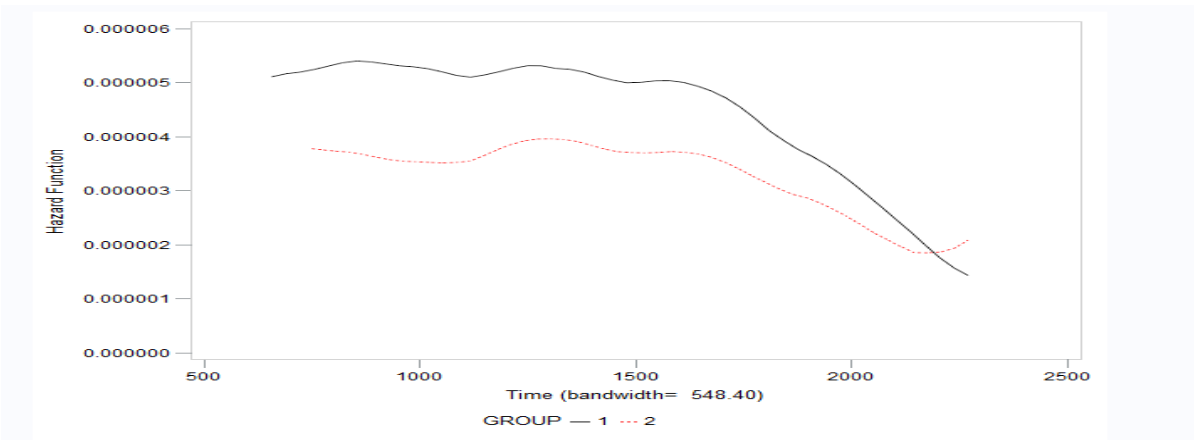

Figure 12. Smooth hazard function

PHREG Model	R-Square	Variables
TIES=DISCRETE	0.71	Overloaded, Normal, Avg_temp_f and Avg _kVA_Rating
TIES=EFRON	0.70	Overloaded, Normal, Avg_temp_f and Avg _kVA_Rating
Stepwise	0.693	Overloaded, Normal, Avg_temp_f and Avg _kVA_Rating
Forward	0.692	Overloaded, Normal, Avg_temp_f and Avg _kVA_Rating
Backward	0.69	Overloaded, Normal, Avg_temp_f and Avg _kVA_Rating

Table 7. PHREG model comparison

Various PHREG models such as Stepwise, Forward, Backward and models with option TIES=EFRON (uses the approximate likelihood of Efron (1977)) and options TIES=DISCRETE (replaces the proportional hazards model by the discrete logistic model,

$$\frac{\mathrm{h}(t;z)}{1-\mathrm{h}(t;z)} = \frac{\mathrm{h}o(t;z)}{1-\mathrm{h}o(t;z)}\exp(z''\beta)$$

Where, h0 (t) and h (t; z) are discrete hazard functions that are built.
The R-Square value is computed using the formula,

$$\mathrm{R}^2 = 1 - \exp(-\frac{G^2}{n})$$

G – Likelihood Ratio and n – Sample size

Using R-Square value as the assessment criterion, we compared the built models and found the PHREG model with TIES=DISCRETE to be the best model for this data set.

MODELING RISK ASSOCIATED WITH OVERLOADING OF TRANSFORMERS

There are two approaches that can be used in modeling risk associated with the overloading of the transformers. The figure below shows the risk categories of transformers based on transformer type and average loaded percentage. The average loaded percentage is the average maximum load calculated across various years.

Type of Transformer	Avg_Loaded_Max	Risk Category
Residential	>=160%	High
	120%<= and <=160%	Medium
	<120%	Low
Commercial	>=120%	High
	100%<= and <=120%	Medium
	<100%	Low

Table 8. Risk categories of transformer

METHOD 1: LOAD BASED LINEAR REGRESSION – WITH AND WITHOUT AGE AS AN INDEPENDENT VARIABLE

In this methodology (Figure 13), the target variable is the continuous variable Average maximum load (Avg_Loaded_Max). The independent variables are as shown in the data table (Table 9). After predicting the load, it is binned into high, medium and low risk categories based on the risk category (Table 8).

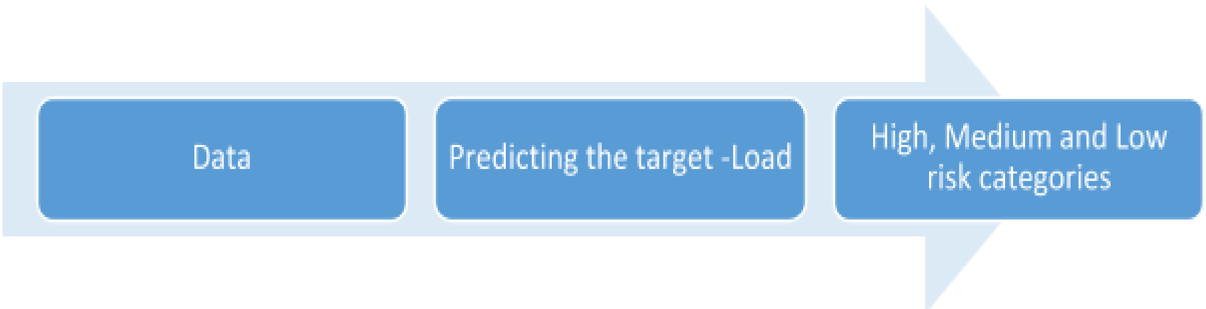

Figure 13. Method 1

S.NO	Variable	Type	Description
1	Avg_pf_max	Continuous	Power Factor of transformers
2	Avg_Temp _f	Continuous	Average maximum temperature of the transformer
3	Avg_Loaded_ Max	Target/ Continuous	Average maximum load of the transformer
4	Age	Continuous	Age of the transformer in days
5	Avg_max_kVA	Continuous	Average kVA rating of the transformer
6	Normal	Continuous	Number of times the transformer was normal
7	Overloaded	Continuous	Number of times the transformer was overloaded
8	Cat_Ind	Binary	Residential - 0 Commercial -1

Table 9. Data – Method 1

There are two different data sets used in this research. One data set covers 2006-2013 transformers where ages of the transformers are known. Therefore regression modeling is performed on these data with age as an input variable. The other data consists of 196,280 transformers where ages of the transformers are not known exactly, hence for these transformers regression modeling is performed without age as an independent variable.

DATA PARTITION

The data from both the data set is split into 40% training, 30% validation and 30% testing.

TRANSFORMATION

Avg_Loaded_Max, Avg_max_kVA, Overloaded have substantial positive skeweness and hence these variables are log transformed. Avg_temp_f and Normal have moderate positive skeweness and hence these variables are square transformed.

MODEL COMPARISON

Model	Age	Training ASE	Validation ASE
Regression (default)	Yes	0.53865	0.546903
Stepwise Regression	Yes	0.53865	0.546871
Forward Regression	Yes	0.53865	0.546871
Backward Regression	Yes	0.53865	0.546871
Decision Tree	No	0.22679	0.23081
Stepwise Regression	No	0.24794	0.24570
Forward Regression	No	0.24794	0.24570
Regression (default model)	No	0.24795	0.24567

Table 10. Model comparison

Models such as full regression, stepwise selection, backward selection, forward selection regression and Decision trees are built with the selection criterion as average square error for the transformers that have age as a parameter. From the above output of the model comparison node (Table 10) we can conclude that the Decision tree performs better with the lowest average square error.

Variable Importance

Obs	NAME	LABEL	NRULES	IMPORTANCE	VIMPORTANCE	RATIO
1	LOG_Avg_max_kVA	Transformed Avg_max_kVA	20	1.0000	1.0000	1.0000
2	cat_ind		5	0.2136	0.2142	1.0028
3	SQR_Avg_temp_max_f	Transformed Avg_temp_max_f	12	0.0990	0.0930	0.9392
4	SQR_Avg_pf_max	Transformed Avg_pf_max	9	0.0766	0.0789	1.0302
5	SQR_Normal	Transformed Normal	1	0.0253	0.0186	0.7360

Figure 14. Decision tree model output

LOG_Avg_max_kVA, Cat_Ind, SQR_Avg_temp_max_f, SQR_Avg_pf_max and SQR_Normal are the important variables contributing to the transformed load factor in the Decision tree model (Figure 14). The predicted Avg_Loaded_max is then binned into appropriate risk categories for residential and commercial transformers (Table 8).

DATA FOR ANALYSIS – METHOD 2

Num.	Variable	Type	Description
1	Avg_pf_max	Continuous	Power Factor of transformers
2	Avg_Temp _f	Continuous	Average maximum temperature of the transformer
4	Age	Continuous	Age of the transformer in days
5	Avg_max_kVA	Continuous	Average kVA rating of the transformer
6	Normal	Continuous	Number of times the transformer was normal
7	Overloaded	Continuous	Number of times the transformer was overloaded
8	Cat_Ind	Binary	Residential – 0 Commercial -1
9	Risk_Category	Target 2/ Category	High, Low and Medium

Table 11. Data – Method 2

METHOD 2: CATEGORY BASED LOGISTIC REGRESSION - WITH AND WITHOUT AGE AS AN INDEPENDENT VARIABLE

In this methodology, the target variable is the categorical variable Risk_Category. The independent variables are as shown in the previous table (Table 11). The load is binned into high, medium and low risk categories based on the risk category (Table 8). After binning them into these 3 categories it is predicted with the help of a logistic regression (Figure 15). As in method 1, two different data sets are used in method 2.

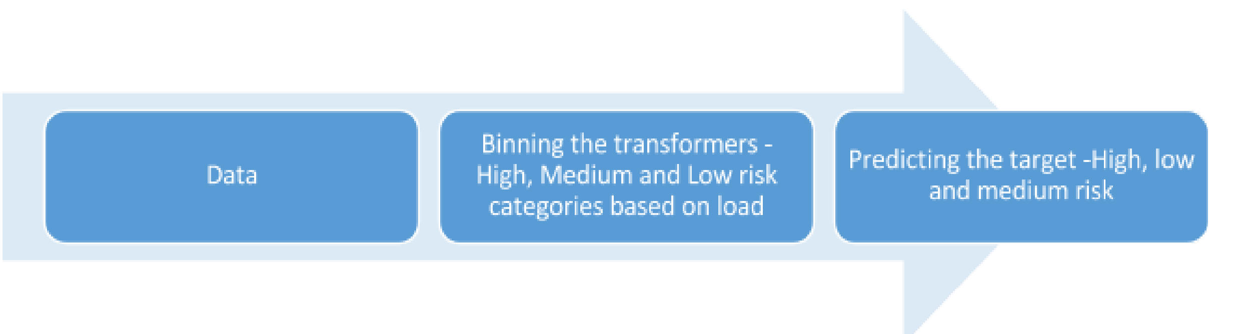

Figure 15. Method 2

DATA PARTITION

The data from both the data set is split into 40% training, 30% validation and 30% testing.

TRANSFORMATION

Avg_Loaded_Max, Avg_max_kVA, Overloaded have substantial positive skeweness and hence these variables are log transformed. Avg_temp_f and Normal have moderate positive skeweness and hence these variables are square transformed.

MODEL COMPARISON

Model	Age	Training Misclassification Rate	Validation Misclassification Rate	Training ASE	Validation ASE
Stepwise Regression	Yes	0.079186	0.078999	0.049572	0.049496
Backward Regression	Yes	0.00001	0.001598	0.000037	0.000739
Forward Regression	No	0.10802	0.10803	0.083926	0.084147
Backward Regression	No	0.10802	0.10803	0.083929	0.084151
Stepwise Regression	No	0.10802	0.10803	0.083929	0.084151
Regression (Default)	No	0.10806	0.10815	0.083935	0.084184

Table 12. Model Comparison Output

Models such as Stepwise regression, Forward regression, Backward regression and Full regressions are built with the selection criterion as Misclassification rate using SAS Enterprise Miner 12.1. From the above output of the model comparison node we can see that the Backward Regression model performs better with the lowest misclassification rate (0.001598) and Average square error (.000739) in the validation data (Table 12).

BACKWARD REGRESSION MODEL RESULTS

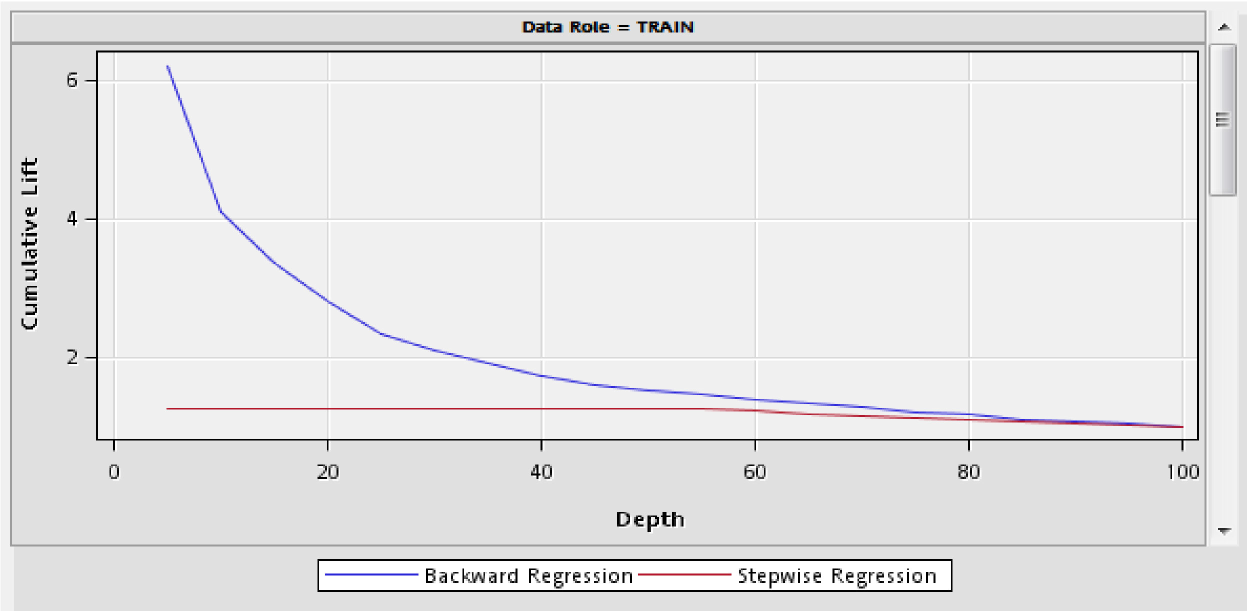

Figure 16. Cumulative Lift Chart

The cumulative lift is higher from a depth of 0 - 60 in Backward Regression compared to Step wise Regression which shows good ability of the model to split the categories efficiently at this depth (Figure 16).

Data Role=VALIDATE Target Variable=Risk_Category Target Label=' '

Target	Outcome	Target Percentage	Outcome Percentage	Frequency Count	Total Percentage
HIGH	HIGH	98.690	100.000	226	4.0121
MEDI	HIGH	1.310	1.370	3	0.0533
LOW	LOW	100.000	99.884	5182	91.9936
LOW	MEDI	2.703	0.116	6	0.1065
MEDI	MEDI	97.297	98.630	216	3.8345

Figure 17. Classification Table Results – Backward Regression

From the classification table results we can see that overall there is 98% accurate prediction in terms of risk categories in the validation data set.

CONCLUSION

The average age of installed transformers in the United States is 40 years. In our analysis we considered short lived transformers (those that were installed after 2006 and failed before 2013). In the short lived transformers, we find that predominantly these fail at the age interval of 1,500-1,625 days. From the PHREG model, we find that the major factors that contribute to the failure rate of transformers in order of their hazard rate are the overloaded count, average temperature, average kVA rating, and the normal condition count of the transformers. We find that the commercial transformers fail faster when compared to the residential transformers. This makes sense considering the harsher conditions under which the commercial transformers operate. Models built to predict the risk factor associated with the transformers suggest that age, average kVA rating and normal condition are the important factors contributing to the overloading of the transformers.

FUTURE SCOPE

An interactive daily dashboard showing high risky transformers can be built so that transformers can be monitored and maintained properly. This can save a lot of cost to the utility companies in terms of repair and damage replacement of equipment. This study concentrates on failing of transformers due to overloading and aging. There can be many other reasons such as failure of insulation material, manufacturing errors, oil contamination, line surge, improper maintenance and lightning which are not considered here. Building a comprehensive model in the future, considering all these factors that impact the failure of transformers will be a great benefactor.

REFERENCES

Yili Hong, William Q. Meeker and James D. McCalley. August 20, 2009. "*Prediction of remaining life of power Transformers based on left truncated and right censored* ". Available at: http://www.researchgate.net/publication/45868253_Prediction_of_remaining_life_of_power_transformers_based_on_left_truncated_and_right_censored_lifetime_data

George Anders - Fellow IEEE, Kinectrics Inc. March 10, 2008. "*Statistical methods and models with analysis of suitability for prediction of the end of life of equipment*". Available at: http://www.pesicc.org/iccwebsite/subcommittees/subcom_a/presentations/spring08/a3.anders.pdf

Paul D. Allison. March 2010. "*Survival Analysis Using SAS®: A Practical Guide, Second Edition*". Available at: http://www.sas.com/store/books/categories/usage-and-reference/survival-analysis-using-sas-a-practical-guide-second-edition/prodBK_61339_en.html

Christian Osorio and Nandan Sawant. *"Transformer Lifetime Prediction, EE292K: Intelligent Energy Projects*". Available at: http://web.stanford.edu/class/ee292k/reports/ChristianNandan.pdf

William H. Bartley P.E., The Hartford Steam Boiler inspection & Insurance Co. "*The Locomotive, Analysis of Transformer Failures, Part 2*". Available at: http://www.hsb.com/TheLocomotive/uploadedFiles/ArticleLibrary/An%20International%20Analysis%20of%20Transformer%20Failures,%20Part%202.pdf

P.K. Sen, PhD, PE, Fellow IEEE, Professor of Engineering, Colorado School of Mines. April 19, 2011. "*Transformer Overloading and Assessment of Loss-of-Life for Liquid-Filled Transformers*." Available at: http://www.pserc.wisc.edu/documents/general_information/presentations/pserc_seminars/psercwebinars20112/Sen_PSERC_Webinar_4-19-11_Slides.pdf

Raymond R. Balise, Health Research and Policy & Spectrum. "*Survival Analysis using SAS Enterprise Guide*". Available at: http://medblog.stanford.edu/lane-faq/survival20100816.pdf

Patricia A. Berglund, Institute For Social Research-University of Michigan, Ann Arbor, Michigan. April 2011. *"An Overview of Survival Analysis using Complex Sample Data".* Available at: http://support.sas.com/resources/papers/proceedings11/338-2011.pdf

Joseph C. Gardiner, Division of Biostatistics, Department of Epidemiology. April 2010. "*Survival Analysis: Overview of Parametric, Nonparametric and Semi parametric approaches and New Developments".* Available at: http://support.sas.com/resources/papers/proceedings10/252-2010.pdf

CONTACT INFORMATION

Your comments and questions are valued and encouraged. Contact the authors at:

Balamurugan Mohan

Graduate Research Assistant- Management Information Systems,

Spears School of Business,

Oklahoma State University, Stillwater, OK – 74075

E-mail: balamurugan.mohan@okstate.edu

Work Phone: 916-768-4559

Balamurugan Mohan is a graduate student in Management Information Systems with SAS® and OSU Data Mining Certificate at Spears School of Business, Oklahoma State University. He has more than 2 and a half years of experience working with Fortune 500 companies. He holds following credentials: SAS® Statistical Business Analyst, SAS® Advanced, Base programmer and SAS® Certified Predictive Modeler.

Dr. Goutam Chakraborty

Oklahoma State University

Stillwater, OK, 74078

E-mail: goutam.chakraborty@okstate.edu

Dr. Goutam Chakraborty is Ralph A. and Peggy A. Brenneman professor of marketing and founder of SAS and OSU data mining certificate and SAS and OSU marketing analytics certificate at Oklahoma State University. He has published in many journals such as Journal of Interactive Marketing, Journal of Advertising Research, Journal of Advertising, Journal of Business Research, etc. He has chaired the national conference for direct marketing educators for 2004 and 2005 and co-chaired M2007 data mining conference. He has over 25 Years of experience in using SAS® for data analysis. He is also a Business Knowledge Series instructor for SAS®.