Estimating the instantaneous survival rate of digital advertising and marketing IDs: LIFESPAN by Cox-Proportional

Nilamadhaba Mohapatra
nilamadhaba.mohapatra@zeotap.com
zeotap
Bangalore, India

Humeil Makhija
humeil.makhija@zeotap.com
zeotap
Bangalore, India

SwapnaSarit Sahu
swapnasarit.sahu@zeotap.com
zeotap
Bangalore, India

ABSTRACT

Finding the active and inactive device IDs[1](ID) in the digital advertising and marketing domain is one of the most crucial tasks in terms of the cost and quality aspect. Keeping the IDs for a longer time will increase the load for the downstream pipelines that incur more storage and computation cost. This can also lead to digital campaigns(advertising or marketing) with low active users thus degrading the performance. Though quality can be improved by putting a constant time to leave(TTL) to each of the IDs, determining an optimal TTL is a tedious task. These IDs are the unique identifiers for the digital domain hence treated as the currency. It also plays an important role in the engineering framework for keeping all other attributes in the storage being linked to it. So, by putting a smaller TTL, losses of ID prematurely can lead to multiple loss of information. This can affect the segment[2] volume export for a campaign largely. On the contrary, if higher TTL is proposed, it can lead to the original problem of cost and computation. Checking an individual ID is active or not in realtime is almost impossible. While most of the non-feedback systems run on TTL based methods to purge the IDs and clean the database, in our paper we propose a granular machine learning-based approach which learns from implicit feedback. We take the bid request from DSPs[3] as feedback which can act as a proxy for individual ID's activity. We created multiple duration parameters from this implicit feedback and experimented with different techniques such as Kaplan-Meier and Cox-Proportional Hazard models to build a robust, learnable, and incremental model. We considered the attributes present for the IDs as covariates and built a Cox Proportional Hazard model with 0.9 concordance score. For a billion scale profile store this is an excellent benchmark.

[1]The device ID is the currency of digital advertising and marketing, whether it be an android ID or an apple ID or any cookie ID.
[2]Segment is a set of device IDs with specific attributes that are used for specific digital advertising or marketing campaigns.
[3]Demand side platform that helps in connecting the media buyers with data exchange platforms.

AdKDD '21, August, 2021, Singapore

ACM ISBN 978-1-4503-XXXX-X/18/06…$15.00

CCS CONCEPTS

• **Computing methodologies** → *Machine learning*; • **Information systems** → **Online advertising**; • **Applied computing** → *Digital cash*; • **Mathematics of computing** → *Survival analysis.*

KEYWORDS

digital advertising, digital marketing, survival analysis, segment quality, digital ID profile store optimization

ACM Reference Format:
Nilamadhaba Mohapatra, Humeil Makhija, and SwapnaSarit Sahu. 2021. Estimating the instantaneous survival rate of digital advertising and marketing IDs: LIFESPAN by Cox-Proportional. In *AdKDD '21: The 27th ACM SIGKDD Conference on Knowledge Discovery and Data Mining, August, 2021, Singapore.* ACM, New York, NY, USA, 6 pages.

1 INTRODUCTION

Online marketing and advertising are never possible without the help of the online identifiers that are associated with electronic devices. It might be the MSDSIN ID for Android, Apple ID for Apple or cookie ID for web browsers. It acts as the primary key for all the attributes and it's properties associated with digital targeting or marketing. Being the currency for the platform, it comes with a great responsibility of storing and managing it properly. People might have multiple devices and IDs associated with the device can change over time implicitly(cookie expiry) or explicitly (resetting the advertisement id by user). So not cleaning the stale IDs periodically from the data store can lead to a significant boost of non-active IDs in the data store. An easy way of solving this problem is to put TTL, but some problems are associated with this kind of solution.

- **Problem 1**: Lack of information regarding the exact time of ID creation makes TTL based methods inefficient. Being a DMP, we might have the time stamp at which the ID enters into our system which is very different from the actual ID creation time in the device.
- **Problem 2** : Each ID with different attributes associated like OS, gender, age, device, bid stream frequency, etc has a different lifetime. In a feedback-based system, we have some uncensored IDs for which we have the lifetime information available. Using this information we can calculate the expected value of life and use it as a TTL. This method still works on the assumption that all the IDs have equal lifetime while the expected value of the uncensored IDs tries to infer the best TTL possible. In this approach, we still have some IDs whose real life is less than the TTL value and we are prolonging their deletion till it reaches the TTL and some IDs whose actual life is more than TTL value and we early

delete it once it reaches the TTL. In both conditions, there is an opportunity to estimate lifetime value more efficiently.

2 BACKGROUND AND RELATED WORK

Overall survival of each group (groups are based on attributes like gender, age, Operating system, etc) can easily be understood using Kaplan-Meier[7]on different groups by analyzing horizontal steps with declining magnitude. However, the detailed understanding of time to event for each device ID can't be understood ignoring the effects of the covariates. Cox Proportional analysis[4] is often used for the detailed analysis of the statistical relationships between a set of covariates and its survival or hazard function.

Historically, survival analysis is mainly used in the medical context where the duration of time until an event is analyzed. It started with the purpose of studying the effects of medicines on the survival of different tuberculosis patients[1]. The idea got extended to the financial markets for predicting the time to the survival of different firms[2, 8], credit card defaulters[5] and time to file business bankruptcy[7] by analyzing the correlation of covariates with the duration parameter (time to event). Survival models with covariates have also been successfully used to analyze the survival of music albums in[3].

Our approach is motivated by[6, 9, 10], where the authors used survival modeling along with covariates to predict the customer attrition rate. We try to extend this idea in the digital advertising and marketing domain to predict the instantaneous survival rate of IDs. We used Kaplan-Meier analysis to understand the behaviour of these IDs belonging to different groups and then used a Cox Proportion-based method to find the instantaneous survival rate of those IDs based on the covariates (i.e. device IDs attributes).

The remainder of the paper is laid out below. In section 3 talks about the survival model analysis techniques. In section 4 and section 5 we discuss the various experiments and results. In Section 6 we talk about the conclusion we made about our approach and propose the future works possible.

3 METHODOLOGY

3.1 Approach

The survival analysis aims to measure the time duration until the event occurs. In our context, we are interested in knowing when a device ID becomes inactive. So the objective for this experiment becomes calculating the survival probability of a device ID at time T after a certain time (t) is defined as:

$$S(t) = Pr(T > t) \tag{1}$$

where T is any non-negative random variable after t where the event may occur.

In this paper we have experimented with both non-parametric (Kaplan–Meier estimator)[7] and semi-parametric (Cox Proportional hazards model)[4] for survival analysis. We started our analysis with the Kaplan-Meier estimator with censored data. It is a univariate and non-parametric method to estimate survival function and it is defined as:

$$S(t) = \prod_{t_i t} \frac{n_i - d_i}{n_i} \tag{2}$$

where d_i is the no of device IDs becomes inactive at time t and n_i is the total number of ids active(censored) at time t .

Kaplan-Meier survival rate estimates of different groups can be understood to have different survival rates. This gives rise to the idea that taking the effect of covariates may give rise to a better estimate of survival probability. We build multiple Cox proportion-based hazard models with different sets of profile attributes as covariates and $MEAN_{HOP}$[4] as a duration parameter.

$$\underbrace{h(t|x)}_{\text{hazard}} = \overbrace{b_0(t)}^{\text{baseline hazard}} \quad \underbrace{\overbrace{\exp\left(\sum_{i=1}^{n} b_i(x_i - \overline{x_i})\right)}^{\text{log-partial hazard}}}_{\text{partial hazard}} \tag{3}$$

where x are the covariates on which hazard rate is dependent on and β is weight for each of the covariates (regression parameters on the log-scale).

The survival function at time t given the covariates X is defined as :

$$\underbrace{S(t|x)}_{\text{survival function}} = \exp\left(- exp\Big(\sum_{i=1}^{n} b_i(x_i - \overline{x_i}) \Big) \int_0^t b_0(u)\mathrm{d}u \right) \tag{4}$$

$$= \exp\left(- exp\Big(\sum_{i=1}^{n} b_i(x_i - \overline{x_i}) \Big) b_0(t) \right) \tag{5}$$

4 EXPERIMENTS

The input data consists of the device ID, their attributes, and the duration parameter. We used our device ID profile store which has around 500M to 10B device IDs depending on the country (like Spain, Italy, India, or the USA) constituting an android advertising ID, IOS advertising ID, and different third parties cookies. Each ID has various static and dynamic attributes like demographic information(Gender, Age, City, State, etc), Application information (APP install and APP usage, etc), device information (device os, device brand, etc), Interest, Intent, and Automotive information. We took random samples from the device ID profile store to get multiple input datasets for our experiment. Here the challenge is to collect the ground truth for the duration parameter for each device ID. We took bid requests from DSPs as implicit feedback as a proxy of device IDs life. The historical bid request data is taken for the last 2 years(**N**) and clubbed **d** days data together into a group. Then we build a Bloom-Filter[5] on top of each group. Overall we have a total of n groups where n is:

$$n = \lfloor \frac{N}{D} \rfloor \tag{6}$$

The i^{th} Bloom-Filter with d days window is defined as β_i^D. Here $N = 600$ days and $D = 10$ days. So, overall we have a timeline of 600 days with 0 referring to the present day. We split our timeline into two parts i.e. 600 to 200 as the experiment set(training set), and 200 to 0 as ground truth for future timesteps prediction. For the ease of experimentation, we further translated the timeline 600

[4]It is defined as the mean time timestamp difference between any two consecutive timestamps at which that device ID was active. Refer Section 4 for details.

[5]Bloom filter is a space-efficient probabilistic data structure often used to inquire whether an element is a member of a set or not.

to 200 to 400 to 0 and predicted for the next 200 days which is 20 timesteps.

Each Bloom-Filter is associated with two timestamps in the the timeline. The upper timestamp defines the the upper timeline boundary(UL) and the lower timestamp defines the lower timeline boundary(LL). For the i^{th} bloom filter they are defined as:

$$UL = d * (i + 1) \tag{7}$$

$$LL = d * i \tag{8}$$

For example a device ID that is found in Bloom Filter$_i$ means it was active at time between $(d * i)$ to $(d * (i + 1))$.

For each device ID, we created a list called $ACTIVE_{LIST}$ containing all the time stamps at which that device ID was active. We have created multiple duration parameters for each device ID out of this data that are defined as below.

- ***ACTIVE$_{LIST}$*** : It is defined as a list containing all the time stamps at which that device ID was active.
- ***ACTIVE$_{MIN}$*** : It is defined as the latest occurrence timestamp of the device ID in $ACTIVE_{LIST}$ (i.e min $ACTIVE_{\text{LIST}}$)
- ***ACTIVE$_{MAX}$*** : It is defined as the first occurrence timestamp of the device ID in $ACTIVE_{LIST}$. (i.e max $ACTIVE_{\text{LIST}}$).
- ***LIFE*** :It defined as the difference between ACTIVE$_{\text{MAX}}$ and $ACTIVE_{MIN}$.

$$LIFE = ACTIVE_{\text{MAX}} - ACTIVE_{\text{MIN}} \tag{9}$$

- ***MAX$_{HOP}$*** : It is defined as maximum timestamp difference in between any two consecutive timestamp of $ACTIVE_{LIST}$.

$$MAX_{\text{HOP}} = max(L_i), \forall L_i \in (ACTIVE_{LIST}(i+1) - ACTIVE_{LIST}(i)), \quad where\ i \in (0\,,\ size(ACTIVE_{LIST}) - 1) \tag{10}$$

- ***MEAN$_{HOP}$*** : It is defined as the mean time timestamp difference between any two consecutive timestamps of $ACTIVE_{LIST}$.

$$MEAN_{\text{HOP}} = \mathbb{E}(L_i), \forall L_i \in (ACTIVE_{LIST}(i+1) - ACTIVE_{LIST}(i)), \quad where\ i \in (0\,,\ size(ACTIVE_{LIST}) - 1) \tag{11}$$

Out of all derived variables we have selected *LIFE*, $MEAN_{HOP}$ and MAX_{HOP} as the candidates chosen for the duration parameter.

In Figure 1 we can observe that the LIFE parameter is not suitable for survival analysis due to its linear decay or very little time-variant nature. Then out of $MEAN_{HOP}$ and MAX_{HOP} we choose $MEAN_{HOP}$ as the best suitable parameter because it is less prone to outliers. A clear explanation to the above point is the case where if an ID is not observed by the bloom filter for a long time and suddenly it's observed by the filter thus increasing its MAX_{HOP} value but it is not the expected(mean) time an ID will take to be seen hence $MEAN_{HOP}$ will not change drastically. So the $MEAN_{HOP}$ parameter can be taken as the best proxy to the duration parameter to calculate the overall survival of an ID.

To understand the survival probability of different groups we plotted the Kaplan-Meier estimate of all groups with the duration parameter $MEAN_{HOP}$ as explained in Figure 2.

Kaplan-Meier survival rate estimates of different groups can be understood to have different survival rates. This provides a

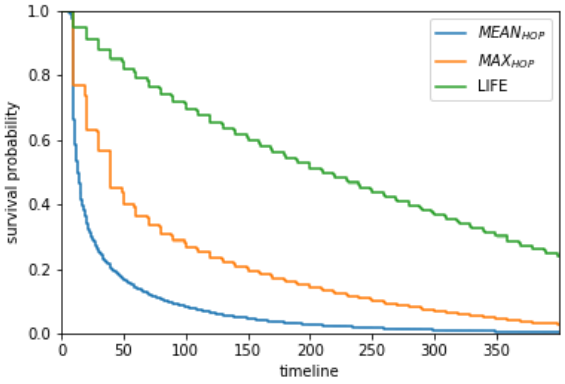

Figure 1: Kaplan–Meier curve for duration parameter importance

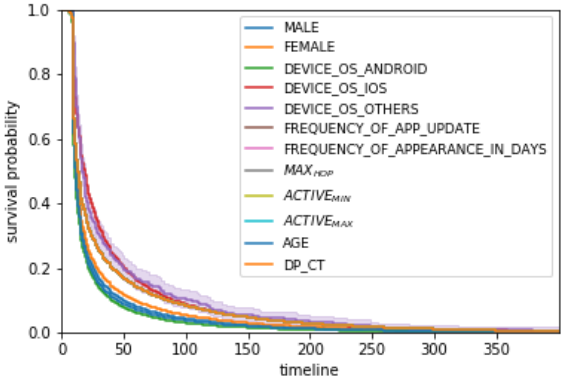

Figure 2: Effect of all the Covariates on MEAN$_{\text{HOP}}$

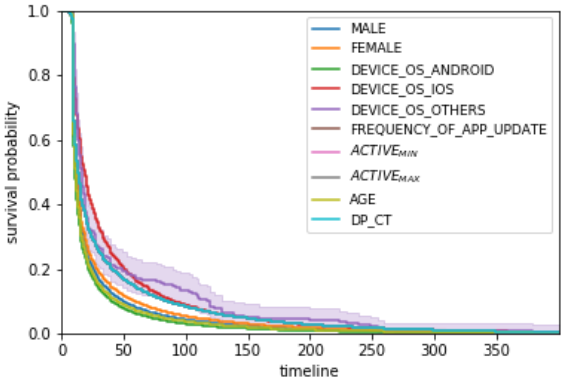

Figure 3: Effect of final Covariates selected on MEAN$_{\text{HOP}}$

notion that covariate based models might be a better estimate of survival probability. We build multiple Cox proportion-based hazard models with different sets of profile attributes as covariates and $MEAN_{HOP}$ as a duration parameter and finally achieved a

Table 1: Effect of covariates on Hazard Rate

COVARIATES	AGE	MALE	FEMALE	DP_CT	DEVICE_OS_ANDROID	DEVICE_OS_IOS	DEVICE_OS_OTHERS	FREQUENCY_OF_APP_UPDATE	ACTIVE MIN	ACTIVE MAX
COEFFICIENT VALUES	-0.00	0.17	0.20	0.10	0.04	0.02	-0.25	-0.01	0.02	-0.02
HAZARD RATIO	0.99	1.18	1.22	1.10	1.04	1.01	0.78	0.99	1.017	0.980

concordance score of 0.9 with the covariates (*AGE*, *DP_,CT*, *FREQUENCY_OF_APP_UPDATE*, *GENDER (MALE, FEMALE)*, *DEVICE_OS(ANDROID, IOS, OTHERS)*) shown in Fig 3.

- **AGE**:Numeric value range between 18 to 100 defined by Demographic age of the device ID.
- **DP_CT**:Numeric value defined by number of data partner contributed to the profile information of the device ID.
- **FREQUENCY_OF_APP_UPDATE**:Numeric attribute defined by frequency (in no. of days) of application data being updated in the device ID profile store.
- **GENDER**:Categorical value defined by demographic gender of the device ID. It can take a value either MALE or FEMALE.
- **DEVICE OS**:Categorical value defined by the device IDs operating system. It can take a value of ANDROID, IOS or others (which includes windows, linux etc).

5 EXPERIMENTAL RESULTS

For the best observed Cox proportion model with the most suitable covariates along with their coefficient values (β) and hazard ratios are shown in Table 1. As we see the covariates *frequency_of_app_update* and *Age* has a little effect on the hazard rate. But IDs with gender as Male or Female tend to suffer from lower survival rates with a higher risk of 18% and 22% respectively. We also observed that the device IDs with the operating system as Android or IOS have a little but negative impact on the hazard rate but the device IDs having operating systems like Blackberry, Windows, Meego, Linux, etc (categorized as others) are associated with a very low hazard rate of -22% thus lives longer. IDs with high value for *DP_CT* found to increase the hazard by 10%. The reason could be more the value of *DP_CT* more the confidence we have about the covariates of the device IDs. We also observed that the last active timestamp tends to be associated with a 1.7% higher hazard rate.

Using the above model we predicted the survival probability for each of the device IDs for the future timesteps. A threshold 0.07 or less is used to flag a device ID as dead and deleted from the storage. We ran the model and compared our results with industry-standard TTLs(of 90, 120, or 180 days) along with the Mean of the LIFE (ground truth). The results and analysis can be found in Table 3. Our analysis is based on 7 main attributes that are defined below :

- **MODEL LIFETIME** ($\boldsymbol{\alpha}$) : Then model predicted lifetime for a device ID.(in number of days)

$$\alpha = T_i^{ID} - ACTIVE_{MAX}^{ID} \tag{12}$$

Table 2: Model stability and robustness check w.r.t sample size

EXP	MODEL AND TTL BASED FEATURES				
	α_μ	γ_μ	$MAE_{\alpha-\gamma}$	99% CI of $MAE_{\alpha-\gamma}$	SAMPLE SIZE
EXP-1	284.15	301.12	49.11	[48.69-49.54]	33,207
EXP-2	287.78	303.46	51.71	[51.52-51.99]	76,855
EXP-3	287.37	303.38	50.75	[50.54-51.81]	316,641
EXP-4	298.96	309.90	51.59	[51.50-51.68]	811,447
EXP-5	299.23	309.96	51.55	[51.50-51.61]	1,894,083

Ti is the time-step at which the predicted survival probability of an ID drops below the threshold.

- **ACTUAL LIFETIME** ($\boldsymbol{\gamma}$): The actual lifetime observed for a device ID.(in number of days) Defined as

$$\gamma = ACTIVE_{MIN}^{ID} - ACTIVE_{MAX}^{ID} \tag{13}$$

- **MODEL LIFETIME MEAN** ($\boldsymbol{\alpha_\mu}$): It is defined as the mean of the lifetime distribution (Using model prediction values) or the expected value of $\boldsymbol{\alpha}$.

$$\alpha_\mu = \mathbb{E}(\alpha) \tag{14}$$

- **PROBABILISTIC MODEL MEAN** ($\boldsymbol{\omega_\mu}$): It is defined as the mean of the lifetime distribution predicted using probabilistic model or the expected value of $\boldsymbol{\omega}$.

$$\omega_\mu = \mathbb{E}(\omega) \tag{15}$$

- **ACTUAL LIFETIME MEAN** ($\boldsymbol{\gamma_\mu}$) : It is defined as the mean of the lifetime distribution (Using Ground Truth values) or the expected value of $\boldsymbol{\gamma}$.

$$\gamma_\mu = \mathbb{E}(\gamma) \tag{16}$$

- **MEAN ABSOLUTE ERROR(MAE**$_{MODEL\_BASED}$**)** ($\boldsymbol{MAE_{\alpha-\gamma}}$): It is defined as the expected value of $|\boldsymbol{\alpha}$ - $\boldsymbol{\gamma}|$

$$MAE_{\alpha-\gamma} = \mathbb{E}|\boldsymbol{\alpha} - \boldsymbol{\gamma}| \tag{17}$$

- **MEAN ABSOLUTE ERROR(MAE**$_{PROBABILISTIC\_MODEL\_BASED}$**)** ($\boldsymbol{MAE_{\omega-\gamma}}$): It is defined as the expected value of $|\boldsymbol{\omega}$ - $\boldsymbol{\gamma}|$

$$MAE_{\omega-\gamma} = \mathbb{E}|\boldsymbol{\omega} - \boldsymbol{\gamma}| \tag{18}$$

- **MEAN ABSOLUTE ERROR (MAE**$_{TTL\_BASED}$**)** ($MAE_{\gamma_\mu-\gamma}$) : It is defined as the expected value of $|\gamma_\mu$ - $\gamma|$

$$MAE_{\gamma_\mu-\gamma} = \mathbb{E}|\gamma_\mu - \gamma| \tag{19}$$

Table 3: Model accuracy comparisons w.r.t TTL based approach

EXP	MODEL V/S TTL BASED APPROACH							
	MAE_{TTL_BASED} (TTL = γ_μ)	$\Downarrow \epsilon_{MAE_{TTL=\gamma_\mu}}$	MAE_{TTL_BASED} (TTL = 90)	$\Downarrow \epsilon_{MAE_{TTL=90}}$	MAE_{TTL_BASED} (TTL = 120)	$\Downarrow \epsilon_{MAE_{TTL=120}}$	MAE_{TTL_BASED} (TTL = 180)	$\Downarrow \epsilon_{MAE_{TTL=180}}$
EXP-1	107.33	54%	219.71	78%	193.24	75%	150.17	67%
EXP-2	107.40	52%	222.553	77%	196.02	74%	152.34	66%
EXP-3	107.484	53%	222.42	77%	195.88	74%	152.23	67%
EXP-4	107.68	52%	222.106	77%	195.58	74%	152.03	66%
EXP-5	106.56	52%	222.160	77%	195.64	74%	152.09	66%

- **CI** :Confidence interval (For model based MAE).
- **MEAN%_ERROR_REDUCTION_PER_ID** $\Downarrow \epsilon_{MAE_{TTL}}$: It is defined as % error reduced by predicting the life of an ID using a model-based approach compared to any feedback or non-feedback based TTL approach.

$$\frac{|MAE_{TTL\_BASED} - MAE_{MODEL\_BASED}|}{MAE_{TTL\_BASED}} \times 100 \quad (20)$$

To understand the stability and efficiency of our model we experimented with various sample sizes and the detailed results can be found in table 2. We found our model prediction to be robust and stable irrespective of any sample size.

In table 3 we see that our Cox Proportional Based Hazard model performs significantly better than any TTL based method. For any industry-standard TTL of (90,120 or 180) days(non-feedback based) our model reduces down the error by approximately 66% to 77% days per device ID. In a Feedback based systems where the life distribution is known if we use the expected lifetime value as a TTL, then also our model reduces the error up to 52% to 54% days per device ID. We performed the above experiments multiple times and calculated the confidence interval of MAE_{MODEL_BASED} using 2 tile T-tests for mean with 25 samples and the 99% CI was found to be 50.19 - 50.99. This above observation as well as the figure 4 proves that our Cox Proportional Based Hazard model performs significantly better and the reduction of error rate due to the model recommendation is stable.

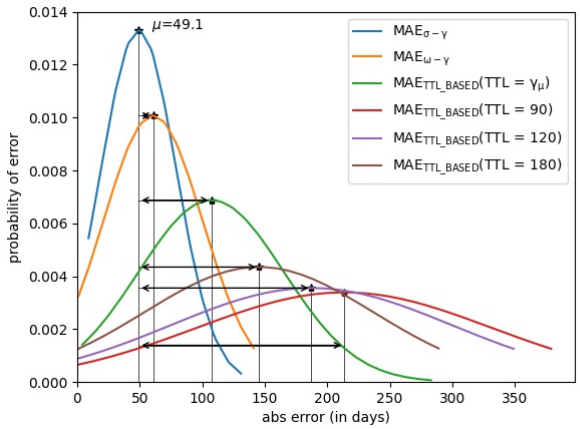

Figure 4: Distribution of MAE_{MODEL_BASED} along with MAE_{TTL_BASED} and $MAE_{PROBABILISTIC\_MODEL\_BASED}$

6 CONCLUSIONS AND FUTURE WORK

The idea of this paper was to find out a technique by which we can decide how long to keep a device ID in our profile store thus optimizing the computation and processing cost. We found out that the covariate based Cox Proportional Hazard model is best suited for this kind of problem. It also provides us a lever to take our decision conservatively (by taking a lower threshold) or proactively (taking a higher threshold) depending on the use case. Our covariate based hazard model achieved a significant concordance score of 0.9. The mean deviation of predicted lifetime of an ID using our model from actual lifetime comes out to be around 50 days from the earlier error rate of 108 days which was using the ground truth mean value as TTL. It narrowed down the average error rate to $\tilde{5}0$ days from 108 days which is 52% more efficient. In terms of actual saving on our billion size data store it helped us to reduce the storage and computation cost by 5% to 8% per run with a frequency of 1 month run. So overall it achieved a savings of around 10% to 16% as the reduction of error rate is approximately 58 days. For a organisation where thousands of downstream workflows are dependent of the data store, this 15% saving in terms on computation and storage is a significant optimisation. However this error rate can be further reduced by using a smaller window bloom filter (3 days or 1 day) instead of 10 days we used above. In future work we would experiment effects of the bloom filter window size on the error rate and find a methodology to find out the best bloom filter size for achieving the optimum result. Our future work is also focused on finding a mechanism to choose the survival threshold more efficiently to ease the manual intervention of decision making.

REFERENCES

[1] Olurotimi Bankole Ajagbe, Zubair Kabair, and Terry O'Connor. 2014. Survival analysis of adult tuberculosis disease. *PloS one* 9, 11 (2014), e112838.
[2] Naohiko Baba, Hiromichi Goko, et al. 2006. Survival analysis of hedge funds. *Institute for Monetary and Economic Studies and Financial Markets Department* 6 (2006).
[3] Sudip Bhattacharjee, Ram D Gopal, Kaveepan Lertwachara, James R Marsden, and Rahul Telang. 2005. The effect of P2P file sharing on music markets: A survival analysis of albums on ranking charts. *Available at SSRN 851284* (2005).

[4] David R Cox. 1972. Regression models and life-tables. *Journal of the Royal Statistical Society: Series B (Methodological)* 34, 2 (1972), 187–202.
[5] Lore Dirick, Gerda Claeskens, and Bart Baesens. 2017. Time to default in credit scoring using survival analysis: a benchmark study. *Journal of the Operational Research Society* 68, 6 (2017), 652–665.
[6] NHE Hasanthika and LALW Jayasekara. 2017. Analyzing the Customer Attrition using Survival Techniques. *International Journal of Statistics and Probability* 6, 6 (2017).
[7] Edward L Kaplan and Paul Meier. 1958. Nonparametric estimation from incomplete observations. *Journal of the American statistical association* 53, 282 (1958), 457–481.
[8] Seigo Matsuno, Yasuo Uchida, Tsutomu Ito, and Takao Ito. 2018. Lifespan of information service firms in Japan: a survival analysis. *IJISPM-INTERNATIONAL JOURNAL OF INFORMATION SYSTEMS AND PROJECT MANAGEMENT* 6, 1 (2018), 61–70.
[9] África Periáñez, Alain Saas, Anna Guitart, and Colin Magne. 2016. Churn prediction in mobile social games: Towards a complete assessment using survival ensembles. In *2016 IEEE International Conference on Data Science and Advanced Analytics (DSAA)*. IEEE, 564–573.
[10] Dirk Van den Poel and Bart Lariviere. 2004. Customer attrition analysis for financial services using proportional hazard models. *European journal of operational research* 157, 1 (2004), 196–217.