

IMPLEMENTASI TD-COX PADA ANALISA SURVIVAL MAHASISWA DI UNIVERSITAS INTERNASIONAL BATAM

Yayuk Setyaning Astutik^{1*}, Dian Tresnawan²

¹Prodi Teknik Sipil, Fakultas Teknik Sipil dan Perencanaan, Universitas Internasional Batam, ² Prodi Teknik Elektro, Fakultas Teknologi Informasi, Universitas Internasional Batam,
JL. Gajah Mada Baloi, Sei-Ladi, Batam Kepulauan Riau 29442.

*E-mail : yayuk@uib.ac.id

Abstrak

Retensi mahasiswa di Universitas telah menjadi perhatian pendidik selama beberapa dekade. Konsekuensi dari kemunduran mahasiswa sangat penting bagi mahasiswa, staf akademik dan Universitas. Dengan demikian, peningkatan retensi mahasiswa merupakan tujuan jangka panjang dari setiap institusi akademik termasuk Universitas Internasional Batam (UIB). Penelitian ini menggunakan data yang disusun dengan melacak 3306 mahasiswa yang terdaftar di UIB mulai tahun 2013 sampai dengan 2017. Variabel dependennya adalah semester pada saat putus kuliah (DO). DO didefinisikan sebagai mahasiswa yang tidak mendaftar dalam semester atau dengan IPK nol. Untuk mengevaluasi kinerja metode yang diusulkan dengan menjalankan dua rangkaian percobaan. Hasil percobaan untuk kedua jenis analisis yaitu prediksi dan estimasi. Membandingkan metode TD-Cox dengan lainnya seperti : Regresi Logistik (LR), Penguatan Adaptif (AB) dan Pohon Keputusan (DT). Kemudian melakukan uji kinerja model untuk memprediksi putus kuliah (DO) pada semester yang berbeda untuk kedua percobaan. Hasil penelitian ini adalah metode TD-Cox merupakan metode yang terbaik dan konsisten. Berdasarkan temuan dari penelitian ini, kita dapat menggunakan informasi pra-pendaftaran seperti tes permulaan untuk mengidentifikasi mahasiswa yang beresiko tinggi untuk putus kuliah. Hal itu juga menunjukkan bahwa satu yang bisa digunakan untuk mengidentifikasi yaitu jumlah yang ditarik atau lulus, status kerja saat kuliah dan IPK setiap semester sebagai peringatan awal untuk melakukan intervensi ketika mahasiswa mempunyai kinerja yang tidak baik.

Kata kunci: Analisis Daya Tahan, TD-Cox, Peningkatan Mahasiswa, Pendaftaran, Prediksi

Abstract

Retention of students at universities has been a concern among educators for many decades. The consequences of student attrition are significant for students, academic staffs and the universities. Thus, increasing student retention is a long term goal of any academic institution included Universitas Internasional Batam. This study using dataset was compiled by tracking 3306 students enrolled at UIB Starting from 2013 until 2017. The dependent variable is the semester of dropout. Dropout is defined as a student who does not register in a semester or whose semester GPA is zero. In order to evaluate the performance of the proposed methods with run two sets of experiment. The experimental result for two types of analysis: predicting dropout student and estimating semester of dropout. Comparing the performance of our proposed TD-Cox and the other methods such as: Logistic Regression (LR), Adaptive Boosting (AB) and Decision Tree (DT). Test the performance of the models to predict the student dropout in different semester for the two experimental setups. The result of this research is the best performance and consistent result of the TD-Cox method. Based on the findings of this study, we can use pre-enrollment information such as screening test to identify student who are at a higher risk of dropping out of their study. It also shows that one can use the number of withdrawn or passed credit, status work during school and GPA at each semester as an early warning to effectively intervene when the students are doing poorly.

Keywords : Survival Analysis, TD-Cox, Student Retention, Enrollment, Prediction.

PENDAHULUAN

Salah satu tujuan jangka panjang dari setiap Universitas adalah mengurangi tingkat kemunduran mahasiswa [2]. Hal tersebut di laporkan bahwa satu dari empat mahasiswa *dropout* dari Universitas setelah tahun pertamanya dan meningkat 50% pada akhir semester keempat [11]. Retensi Mahasiswa merupakan area penelitian yang banyak diteliti di Institusi Pendidikan Tinggi. Banyak institusi yang berkonsultasi pada lembaga konsultan terkait masalah tersebut. Pada dekade sebelumnya, model komprehensif telah dikembangkan untuk mengatasi hal itu secara teori. Kemudian digunakan metode statistika untuk memprediksi *student dropout* dan juga untuk menemukan faktor penting yang mempengaruhi masalah tersebut. Regresi merupakan teknik utama yang bisa diaplikasi pada area ini [10]. Regresi Logistik merupakan metode statistik lainnya yang sering digunakan dalam kasus ini.

Banyak penjelasan yang bisa ditemukan untuk membantu institusi pendidikan dalam memprediksi risiko mahasiswa [6]. Metode tradisional yang bisa digunakan seperti regresi dan regresi logistik yang telah digunakan untuk mengidentifikasi *dropout student* pada beberapa dekade [5]. Sekarang ini, masalah retensi mahasiswa telah menarik perhatian para peneliti khususnya data mining dan merupakan proses yang panjang serta bergantung pada waktu [4]. Sehingga teknik analisis data longitudinal sesuai untuk kondisi tersebut.

Ketika analisis prediksi telah digunakan pada bidang industri selama bertahun-tahun, institusi pendidikan relatif terlambat mengadopsi pendekatan tersebut untuk mendukung pengambilan keputusan.

Begitu juga dengan Universitas Internasional Batam, dimana setiap tahunnya bisa dilihat jumlah mahasiswa yang tidak berimbang antara yang masuk dan lulus. Meskipun 80% mahasiswa menyelesaikan studinya dalam waktu < 8 semester. Tetapi masih ada mahasiswa yang menyelesaikan studinya lebih dari 8 semester. Atas dasar itulah dikembangkan penelitian dengan analisis survival berdasarkan kerangka informasi atribut yang ditetapkan. Untuk tujuan tersebut akan diimplementasikan *Cox* dan *Time-Dependent Cox* (TD-Cox) untuk masalah retensi mahasiswa. Ide dasarnya

adalah menggunakan metode analisis survival pada atribut informasi saat seleksi masuk universitas untuk memprediksi *student dropout*. Kemudian membandingkan dengan metode tradisional lainnya.

KERANGKA TEORI

Dengan mengembangkan *Time-Dependent Model* guna memprediksi *Student Dropout* dengan informasi Seleksi Masuk Universitas dan Setelah Masuk Universitas. Berikut adalah notasi yang digunakan dalam penulisan yang disajikan dalam Tabel 1.

Tabel 1. Notasi yang digunakan dalam penelitian

Notasi	Deskripsi
n	Jumlah Data
p	Jumlah atribut statik
q	Jumlah <i>Time Dependent</i>
$X_i \rightarrow 1 \times p$	Matrix atribut untuk mahasiswa i
$Z_i(t) \rightarrow 1 \times q$	Matrix <i>Time-Dependent</i> untuk mahasiswa i
$Y \rightarrow n \times 1$	Vektor waktu kejadian aktual
$C \rightarrow n \times 1$	Vektor waktu tersensor
$T \rightarrow n \times 1$	Vektor dari waktu pengamatan yaitu $\min(Y, C)$
$\delta \rightarrow n \times 1$	Vektor binary dari status yang tersensor
d_i	Jumlah kejadian waktu t_i
$S_0(t)$	Survival Probability Dasar
$S(t X, Z(t))$	Probabilitas kondisional Survival pada waktu t
$h_0(t)$	Tingkat Hazard Dasar
$h(t X, Z(t))$	Probabilitas kondisional hazard
β	Vektor koefisien dari Regresi Cox
$L(\beta)$	Fs. Maksimum Likelihood untuk β

Kemudian kita akan mendefisikan beberapa kondisi yang digunakan dalam penelitian ini.

- *Dropout Student*: didefinisikan mahasiswa yang tidak register pada semester tersebut atau yang mempunyai IPK semester nol
- *Event: Student dropout* sebelum kelulusan atau kejadian yang menarik.

- *Censored*: Jika mahasiswa tidak *dropout* pada 8 semester pertama atau batas yang ditentukan dalam penelitian.

ANALISIS SURVIVAL

Analisis survival telah menjadi alat penting untuk menganalisis data waktu antar kejadian (*time to event data*) atau menganalisis data yang berhubungan dengan waktu, mulai dari *time origin* sampai terjadinya suatu peristiwa khusus. Pada aplikasi analisis survival, sangat umum terjadi bahwa periode observasi yang menarik adalah tidak lengkap untuk beberapa observasi dan data tersebut dikatakan tersensor.

Misal $D_n(t) = \{X_i, Z_i(t), T_i, \delta_i(t); i=1, \dots, n\}$ menyatakan sampel dari dataset D pada waktu t , dimana X_i menyatakan sebuah vektor kovariate ($1 \times p$) untuk subjek i ketika ada p variabel statistik pada data. $Z_i(t)$ merepresentasikan ($1 \times q$) vektor kovariat dari *time dependent* pada waktu t dan T_i dinyatakan pada waktu kejadian yang terobservasi. Misalkan Y_i adalah *time survival*, tetapi mungkin ini tidak teramati maka dengan mengamati $T_i = \min(Y_i, C_i)$ dimana C_i adalah waktu tersensor. Jika data telah tersensor dan bersama dengan Y_i maka akan diperoleh indikator variabel $\delta_i = \begin{cases} 1 & Y_i \leq C_i \\ 0 & Y_i > C_i \end{cases}$. Jadi untuk

individu i , jika $\delta_i = 0$ artinya tersensor dan jika $\delta_i = 1$ maka tidak tersensor.

Mengingat T adalah durasi menjadi variabel random kontinu fungsi survival, $S(t)$ adalah probabilitas dimana waktu kejadian lebih terlambat dari waktu tertentu t , maka dapat didefinisikan sebagai

$$S(t) = \Pr(T > t) = \int_t^{\infty} f(u) du = 1 - F(t) \quad (1)$$

dimana $f(u)$ adalah fungsi kepadatan peluang dan $F(t)$ adalah fungsi distribusi kumulatif. Fungsi hazard atau tingkat kejadian didefinisikan sebagai:

$$h(t) = \lim_{dt \rightarrow 0} \frac{\Pr(t \leq T < t + dt)}{dt} \quad (2)$$

COX PROPORTIONAL HAZARD

Salah satu metode yang populer pada analisis survival adalah *Cox Proportional Hazard Model* [6]. Model ini merupakan *semi-parametric* dengan sedikit asumsi daripada metode parametrik [4]. *Model Cox* menyajikan secara mudah untuk menginterpretasikan informasi terkait hubungan fungsi hazard ke prediktor. Fungsi Hazard untuk *Cox Proportional Hazard Model* mempunyai bentuk:

$$h(t|X) = h_0(t) \exp \sum_{y=1}^m \beta_y X_y = h_0(t) e^{(\beta X)} \quad (3)$$

dimana $h_0(t) = e^{\alpha(t)}$ adalah *baseline* fungsi

hazard pada waktu t dan $\exp \sum_{i=1}^n \beta_i X_i$ adalah risiko yang terkait dengan nilai kovariat. Oleh karena itu, fungsi probabilitas untuk *Cox Model* dapat dinyatakan sebagai

$$S(t|X) = S_0(t) \exp(\beta X) \quad (4)$$

$$\text{dimana } S_0(t) = e^{-\int_0^t h_0(x) dx} \quad (5)$$

Parameter dari Model Regresi Cox diestimasi dengan meminimalisasi *Partial Likelihood* [7]. Berdasarkan formula Regresi Cox, *partial likelihood* bisa dikonstruksikan

$$\text{dari dataset bentuk: } L(\beta) = \prod_{i: \delta_i=1} \frac{\theta_i}{\sum_{j: t_j \geq t_i} \theta_j} \quad (6)$$

dimana $\theta_i = \exp(\beta X_i)$ dan (X_1, \dots, X_n) adalah vektor kovariat untuk n individu sampel independen pada dataset. Dengan menyelesaikan $\frac{\partial L(\beta)}{\partial \beta} = 0$, koefisien kovariat

dapat diestimasi dengan $\hat{\beta}$. Untuk mendapatkan *baseline fungsi hazard* pada fungsi *likelihood*, dimana β dapat diganti dengan $\hat{\beta}$. Kemudian $h_0(t_i)$ dapat diperoleh

$$\text{dari } \hat{h}_0(t_{(i)}) = 1 / \sum_{j \in R(t_{(i)})} \theta_j \quad (7)$$

TIME-DEPENDENT COX (TD-COX)

Regresi *Cox Proportional Hazard* mempunyai asumsi bahwa kovariat tidak tergantung pada waktu. Dengan kata lain, ketika kovariat tidak berubah dari waktu ke

waktu atau saat data hanya dikumpulkan untuk kovariat pada satu titik waktu, maka metode yang tepat adalah menggunakan variabel statis untuk menjelaskan hasilnya.

Di sisi lain dalam banyak situasi (seperti masalah retensi mahasiswa), maka kita bisa mendefinisikan variabel *time-dependent* yang bisa berubah nilainya pada waktu pengamatan. Untuk lebih memahaminya diberikan contoh ilustrasi. Tabel 2 menunjukkan format data per mahasiswa. Pada contoh ini, untuk setiap mahasiswa dicatat *Time Dropout*, Status dan IPK yang diperoleh Setelah Masuk Kuliah. Untuk Status 1 artinya *Student Dropout* dan untuk 0 *Student Not Dropout* sampai waktu observasi.

Tabel 2. Contoh Data Mahasiswa (IPK (1) terhadap IPK Semester Pertama)

Stud ent	Time	Status	IPK(1)	IPK(2)	IPK(3)
ID_1	1	1	2	-	-
ID_2	2	1	3.2	1.8	-
ID_3	3	0	4	4	3.5

Kemudian, untuk *time-dependent* analisis survival membutuhkan format untuk menghitung prosesnya dengan format yang berbeda dan disajikan dalam Tabel 3.

Tabel 3. Contoh Data Survival Berdasarkan Formatting

Student	t_0	t	Status	IPK
ID_1	0	1	1	2
ID_2	0	1	0	3.2
ID_2	1	2	1	1.8
ID_3	0	1	0	4
ID_3	1	2	0	4
ID_3	2	3	0	3.5

Pada penelitian ini, akan dikembangkan *Time Dependent Cox Regression* yang dikenal *TD-Cox*, yang secara simultan akan menangani kovariat statis dan tergantung waktu. Kemudian Fungsi Hazard-nya bisa didefinisikan sebagai:

$$h(t|X, Z(t)) = h_0(t) e^{\beta(X+Z(t))} \quad (8)$$

Konsekuensinya, fungsi probabilitas survival untuk *TD-Cox Model* dapat diformulasikan sebagai:

$$S(t|X, Z(t)) = S_0(t)^{\exp(\beta(X+Z(t)))} \quad (9)$$

Perlu diketahui bahwa *TD-Cox* Parameter dapat diperoleh dari simulasi data berdasarkan fungsi *Maksimum Likelihood*. Kemudian dengan menggunakan (9), dapat diestimasi

peluang *dropout* dari setiap mahasiswa pada data tes.

ALGORITMA : TD-Cox Method

Input : Student data $D_n(t) = \{X, Z(t), T, \delta\}$

Output: Probability of Student Dropout

Bagian 1 : Reformatting Data

for $i=1$ to n **do**

$T_c \leftarrow T_i$;

for $j=0$ to T_c **do**

for $k=1$ to q **do**

$Z_k = Z_k(j)$;

$t_{i+j} = j$;

if $\delta_i = 1$ and $j = T_c$ **then**

$S_{i+j} = 1$;

else

$S_{i+j} = 0$

end

end

end

end

Bagian 2 : TD-Cox Method

TD-Cox Method, β_s dan \hat{h}_0 menggunakan persamaan (6) dan (7).

For each student in the test data do

Estimate $\hat{S}(t|X, Z)$ from Eq. (9)

$\hat{F}(t|X, Z) = 1 - \hat{S}(t|X, Z)$

end

METODE

Dalam penelitian ini, data set disusun dengan melacak 3306 mahasiswa yang terdaftar di Universitas Internasional Batam mulai angkatan tahun 2013 sampai dengan 2017. Diantaranya ada yang *dropout* pada akhir tahun pertama, kedua, ketiga dan keempat. Dependen variabel adalah semester dropout. Dalam penelitian ini, dropout didefinisikan sebagai mahasiswa yang tidak mendaftar atau yang mempunyai IPK semester nol. Untuk mengevaluasinya, kita menggunakan dua rangkaian percobaan dengan ketentuan sebagai berikut: (1) Mengumpulkan informasi mahasiswa yang masuk di UIB dari 2013 sampai 2017 dan

mencatat rekaman mahasiswa sampai 8 semester dan (2) Memotong observasi di semester Ganjil 2017, karena disemester ini hanya memiliki catatan satu semester. Setelah persiapan data dan *screening* data, akhirnya diperoleh 21 atribut prediktor dengan kategori yaitu: Demografi, Latar Belakang Keluarga, Keuangan, Informasi Asal Sekolah, Asal Daerah, Status Pekerjaan Saat Kuliah, Seleksi Masuk Universitas dan Setelah Masuk Universitas. Dengan menggunakan perbandingan Metode *TD-Cox* dengan metode tradisional lainnya (*Logistic Regression-LR*, *Adaptive Boosting-AB*, *Decision Tree-DT*, *Cox* dan *Time Dependent Cox-TD Cox*).

HASIL DAN PEMBAHASAN

Pada bagian ini disajikan hasil analisis survival mahasiswa untuk memprediksi *dropout*. Menjelaskan sumber data yang digunakan, menentukan variabelnya untuk modelnya dan menggambarkan evaluasi metrik yang digunakan untuk membandingkan hasil yang diusulkan dengan metode lainnya.

Daftar lengkap atribut dan deskripsi disajikan dalam Tabel 4 untuk digunakan analisa perbandingan antara metode *TD-Cox* dengan lainnya.

Tabel 4. List Atribut yang digunakan dalam penelitian.

List Atribut	
Demografi :	Informasi Asal Sekolah:
- Jenis Kelamin	- Jenis Sekolah Asal
- Suku	- Nilai UN
- Agama	- Asal Sekolah
- Tempat Tinggal	Status Pekerjaan Selama Kuliah :
Latar Belakang Keluarga :	- Kuliah Dengan Kerja
- Jenjang Pendidikan Ayah	- Jenis Pekerjaan
- Jenjang Pendidikan Ibu	Seleksi Masuk Universitas:
- Jumlah Anggota Keluarga yang menempuh pendidikan di Universitas	- Gelombang Pendaftaran/Seleksi
Keuangan:	- Jenis Mahasiswa Beasiswa/Tidak
- Jumlah Pengeluaran Mahasiswa	Setelah Masuk Universitas:
- Jumlah Pengeluaran Orang Tua	- Kelas Pagi/Sore
- Pendapatan Mahasiswa	- Persentase SKS lulus
- Pendapatan Orang Tua	- Persentase SKS gagal
	- IPK Semester

Agar memiliki ukuran kuantitatif untuk mengestimasi model yang diusulkan dengan membandingkan teknik klasifikasi yang lain, maka digunakan dua set percobaan. Membagi data set ke dalam "Training" dan "Testing". *Training* terdiri dari data yang terdiri dari

catatan mahasiswa yang sudah diterima dari tahun 2013 sampai dengan 2016.

Data *Testing* yang terdiri dari mahasiswa yang diterima pada tahun 2017 tidak terpakai untuk membangun model. Validasi dilakukan sebanyak 10-fold pada data *Trainng* dan *Testing* pada tabel yang terpisah. Dengan menggunakan teknik *Cross Validation Stratified* sebanyak 10 kali, dimana membagi setiap data set menjadi 10 subset, yang disebut *fold* dengan kriteria yang hampir sama dan distribusi yang sama dari mahasiswa yang *dropout* dan *non-dropout*. Pada setiap percobaan, satu kali *fold* digunakan untuk menguji model yang telah dikembangkan dari 9 *fold* yang tersisa selama fase *Training*. Kemudian setiap metode dihitung sebagai rata-rata dari 10 percobaan. Metode ini kita terapkan dalam bahasa pemrograman R dengan menggunakan *package* "Survival". Untuk menilai model yang diusulkan dalam masalah klasifikasi maka digunakan metrik berikut ini:

- (1) *Accuracy* dinyatakan dalam bentuk presentase subjek pada *Testing* yang diklasifikasikan dengan benar.
- (2) *F-Measure* didefinisikan sebagai *Mean of Precision* dan *Recall* yang harmonis. Nilai *F-Measure* yang tinggi menunjukkan bahwa *Precision* dan *Recall* juga cukup tinggi.

$$F - Measure = \frac{2 \times Precision \times Recall}{Precision + Recall}$$

$$Precision = \frac{TP}{TP + FP} \text{ dan } Recall = \frac{TP}{TP + FN}$$

dimana TP : *True Positive*, FP : *False Positive* dan FN : *False Negative*.

- (3) *AUC* dinyatakan sebagai area dibawah ROC (*Reciever Operating Characteristic*) dimana kurva dibuat dengan membuat plot TPR (*True Positive Rate*) terhadap FPR (*False Positive Rate*) dengan berbagai nilai *threshold*.

Untuk memprediksi *Time to Dropout* digunakan *Mean Absolute Error* (MAE).

- (4) *Mean Absolute Error* (MAE) adalah Jumlahan yang digunakan untuk mengukur seberapa dekat hasil prediksi terhadap hasil aktual yang dinyatakan dalam bentuk:

$$MAE = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i|$$

dimana \hat{y}_i adalah nilai prediksi dan y_i adalah nilai sebenarnya untuk subjek i .

Untuk MAE juga dilakukan evaluasi model yaitu :

- (1) *Underestimated Prediction Error Rate* (UPER) yang didefinisikan sebagai pecahan dari Underestimated Prediction Output atas Prediction Error keseluruhan.

$$UPER = \frac{\sum_{i=1}^n I(\hat{y}_i < y_i)}{\sum_{i=1}^n I(\hat{y}_i < y_i) + \sum_{i=1}^n I(\hat{y}_i > y_i)}$$

- (2) *Overestimated Prediction Error Rate* (OPER) : kesalahan total adalah konstan, sehingga OPER dapat dinyatakan dalam bentuk:

$$OPER = 1 - UPER$$

Hasil percobaan untuk dua jenis analisis yaitu *Predicting Dropout Student* dan *Estimating Semester of Dropout*.

(a) *Predicting Dropout Student*

Dengan membandingkan hasil dari *TD-Cox* dan metode *Standar Cox* terhadap ketiga teknik klasifikasi yaitu *Logistic Regression* (LR), *Adaptive Boosting* (AB), dan *Decision Tree* (DT). Kita menguji hasil dari model untuk memprediksi *Student Dropout* disemester yang berbeda dari percobaan yang didesain seperti dijelaskan diatas. Hasilnya ditunjukkan pada Tabel 5-8 . Dari Tabel tersebut, kita dapat melihat bahwa hasil yang terbaik dan hasil yang konsisten diperoleh dari metode TD-Cox. Pada penelitian ini, seperti yang dijelaskan pada Tabel 4, maka didefinisikan 4 variabel Setelah Masuk Universitas termasuk Kelas Pagi/Sore, Persentase SKS yang Lulus, Persentase SKS yang Gagal dan IPK Semester. Ketika menggunakan atribut tersebut bersama dengan *Seleksi Masuk Universitas* pada *TD-Cox* model

yang diusulkan, maka akan diperoleh hasil klasifikasi yang lebih baik. Kemudian, tidak seperti metode klasifikasi lainnya, *TD-Cox* mempunyai kemampuan menggunakan informasi yang lebih pada atribut *Setelah Masuk Universitas* dengan memperkenalkan *time dependent* pada model. Disisi lain, walaupun menggunakan atribut *Seleksi Masuk Universitas*, *Cox* menyajikan hasil yang lebih baik jika dibandingkan dengan metode klasifikasi yang lain. Berdasarkan hasil tersebut, metode Cox meningkatkan *Prediction Accuracy of Student Dropout* sekitar 8.7%. Sehingga Model Regresi Cox akan menjadi pilihan yang lebih baik untuk masalah klasifikasi dalam Longitudinal data jika dibandingkan dengan metode tradisional.

(b) *Estimating Semester of Dropout*

Tabel 13 menunjukkan hasil *10-fold cross validation* data dan 2017 sebagai data tes dengan menggunakan rancangan *experimental* 1 dan 2. Kemudian membandingkan hasil Cox dengan *Linier Regression* dan *Support Vector Regression* (SVR). *TD-Cox* tidak bisa digunakan untuk tujuan ini dan hanya bisa digunakan untuk informasi *Seleksi Masuk Universitas* guna mengestimasi *Semester of Dropout*. *TD-Cox* bisa digunakan sebagai informasi Setelah Masuk Kuliah setelah semester kuliah dimulai. Berdasarkan Tabel 9 dengan adanya data yang tersensor, untuk analisis survival, Cox model lebih baik dari metode regresi tradisional. Hal ini juga bisa dilihat Cox mempunyai nilai UPER tinggi, hal ini mengindikasikan bahwa *error* secara keseluruhan berasal dari *underestimating the semester of dropout*.

Kemudian, dapat dilihat bahwa nilai standar deviasi untuk LR, AB, DT, Cox dan TD-Cox baik untuk Experimen 1 dan Experimen 2 yang disajikan dari Tabel 9-12.

Tabel 5. Hasil dari LR, AB, DT dengan Cox dan TD-Cox Retensi Mahasiswa di UIB dari 2013-2016 (Exp 1) untuk Setiap Semester dengan *10-fold Cross Validation*.

Semester	Accuracy					F-Measure					AUC				
	LR	AB	DT	Cox	TD-Cox	LR	AB	DT	Cox	TD-Cox	LR	AB	DT	Cox	TD-Cox
1	0.715	0.719	0.716	0.729	0.729	0.712	0.722	0.710	0.734	0.734	0.744	0.757	0.717	0.751	0.751
2	0.741	0.747	0.735	0.752	0.765	0.716	0.739	0.729	0.748	0.751	0.782	0.783	0.734	0.786	0.792
3	0.761	0.767	0.744	0.779	0.887	0.744	0.750	0.733	0.759	0.762	0.789	0.798	0.741	0.802	0.816
4	0.782	0.778	0.751	0.781	0.801	0.760	0.761	0.739	0.775	0.784	0.792	0.814	0.754	0.812	0.829
5	0.783	0.775	0.753	0.792	0.812	0.772	0.769	0.745	0.789	0.792	0.811	0.819	0.768	0.825	0.837

6	0.798	0.789	0.775	0.986	0.821	0.782	0.776	0.753	0.797	0.805	0.823	0.830	0.779	0.831	0.844
7	0.821	0.814	0.802	0.803	0.839	0.791	0.798	0.768	0.802	0.819	0.831	0.841	0.802	0.840	0.851
8	0.857	0.849	0.837	0.818	0.841	0.806	0.818	0.775	0.819	0.828	0.844	0.852	0.819	0.852	0.860

Tabel 6. Hasil dari LR, AB, DT dengan *Cox* dan TD-*Cox* Retensi Mahasiswa di UIB pada 2017 (Exp1) untuk Setiap Semester.

Semester	Accuracy					F-Measure					AUC				
	LR	AB	DT	Cox	TD-Cox	LR	AB	DT	Cox	TD-Cox	LR	AB	DT	Cox	TD-Cox
1	0.711	0.719	0.698	0.715	0.715	0.713	0.721	0.697	0.719	0.719	0.726	0.733	0.711	0.742	0.742
2	0.737	0.741	0.727	0.739	0.745	0.731	0.732	0.723	0.732	0.745	0.753	0.779	0.735	0.783	0.787
3	0.757	0.755	0.739	0.757	0.773	0.757	0.750	0.735	0.751	0.768	0.789	0.785	0.739	0.797	0.805
4	0.778	0.771	0.747	0.778	0.784	0.769	0.764	0.742	0.765	0.789	0.827	0.829	0.758	0.824	0.828
5	0.789	0.780	0.758	0.785	0.805	0.781	0.776	0.749	0.789	0.802	0.834	0.829	0.770	0.830	0.839
6	0.791	0.788	0.779	0.799	0.830	0.788	0.789	0.771	0.796	0.816	0.842	0.839	0.783	0.836	0.844
7	0.802	0.819	0.826	0.836	0.841	0.794	0.812	0.825	0.833	0.829	0.856	0.841	0.791	0.841	0.851
8	0.816	0.825	0.834	0.849	0.609	0.805	0.809	0.819	0.827	0.841	0.866	0.858	0.810	0.859	0.867

Tabel 7. Hasil dari LR, AB, DT dengan *Cox* dan TD-*Cox* Retensi Mahasiswa di UIB dari 2013-2016 (Exp 2) untuk Setiap Semester dengan 10-fold Cross Validation.

Semester	Accuracy					F-Measure					AUC				
	LR	AB	DT	Cox	TD-Cox	LR	AB	DT	Cox	TD-Cox	LR	AB	DT	Cox	TD-Cox
1	0.705	0.709	0.716	0.719	0.719	0.702	0.712	0.701	0.724	0.724	0.734	0.747	0.707	0.751	0.751
2	0.731	0.737	0.735	0.742	0.751	0.726	0.728	0.719	0.738	0.741	0.772	0.783	0.724	0.786	0.792
3	0.739	0.743	0.749	0.757	0.768	0.734	0.739	0.723	0.749	0.752	0.789	0.798	0.731	0.805	0.816
4	0.748	0.753	0.751	0.767	0.777	0.750	0.751	0.739	0.765	0.774	0.811	0.814	0.754	0.818	0.828
5	0.767	0.762	0.769	0.785	0.796	0.762	0.757	0.745	0.779	0.790	0.817	0.829	0.758	0.825	0.838
6	0.778	0.770	0.789	0.792	0.814	0.773	0.766	0.751	0.787	0.811	0.827	0.831	0.769	0.831	0.840
7	0.781	0.793	0.811	0.823	0.841	0.781	0.789	0.763	0.791	0.833	0.830	0.842	0.777	0.841	0.849
8	0.795	0.802	0.820	0.839	0.850	0.790	0.811	0.781	0.800	0.849	0.847	0.857	0.789	0.850	0.861

Tabel 8. Hasil dari LR, AB, DT dengan *Cox* dan TD-*Cox* Retensi Mahasiswa di UIB pada 2017 (Exp2) untuk Setiap Semester.

Semester	Accuracy					F-Measure					AUC				
	LR	AB	DT	Cox	TD-Cox	LR	AB	DT	Cox	TD-Cox	LR	AB	DT	Cox	TD-Cox
1	0.711	0.719	0.698	0.715	0.715	0.713	0.710	0.697	0.719	0.719	0.716	0.723	0.701	0.742	0.742
2	0.737	0.741	0.717	0.739	0.745	0.721	0.723	0.711	0.732	0.745	0.743	0.749	0.725	0.773	0.787
3	0.745	0.750	0.725	0.753	0.767	0.739	0.734	0.725	0.749	0.768	0.768	0.762	0.728	0.789	0.801
4	0.762	0.754	0.738	0.764	0.774	0.747	0.743	0.731	0.768	0.789	0.793	0.778	0.735	0.791	0.820
5	0.778	0.772	0.749	0.781	0.792	0.756	0.759	0.742	0.779	0.809	0.813	0.798	0.749	0.816	0.835
6	0.784	0.779	0.758	0.797	0.820	0.765	0.773	0.754	0.787	0.829	0.825	0.805	0.758	0.827	0.839
7	0.791	0.788	0.777	0.811	0.831	0.778	0.780	0.767	0.790	0.833	0.839	0.819	0.768	0.830	0.846
8	0.800	0.799	0.790	0.828	0.844	0.790	0.791	0.799	0.818	0.841	0.842	0.823	0.777	0.841	0.851

Tabel 9. Standar Deviasi dari LR, AB, DT dengan *Cox* dan TD-*Cox* Retensi Mahasiswa di UIB dari 2013-2016 (Exp 1) untuk Setiap Semester dengan 10-fold Cross Validation.

Semester	Accuracy					F-Measure					AUC				
	LR	AB	DT	Cox	TD-Cox	LR	AB	DT	Cox	TD-Cox	LR	AB	DT	Cox	TD-Cox
1	0.023	0.019	0.035	0.015	0.015	0.031	0.028	0.045	0.029	0.029	0.015	0.013	0.026	0.013	0.013
2	0.025	0.023	0.037	0.019	0.018	0.033	0.029	0.051	0.031	0.030	0.018	0.015	0.023	0.013	0.011
3	0.024	0.019	0.034	0.018	0.016	0.034	0.033	0.048	0.028	0.029	0.017	0.016	0.019	0.013	0.012
4	0.025	0.027	0.031	0.017	0.018	0.036	0.032	0.038	0.029	0.027	0.016	0.018	0.021	0.014	0.013
5	0.025	0.029	0.039	0.020	0.019	0.034	0.031	0.039	0.027	0.026	0.016	0.015	0.019	0.012	0.012
6	0.028	0.024	0.031	0.017	0.015	0.030	0.035	0.043	0.028	0.024	0.012	0.015	0.021	0.011	0.009
7	0.025	0.026	0.034	0.018	0.016	0.032	0.034	0.038	0.028	0.026	0.014	0.014	0.020	0.012	0.011
8	0.026	0.025	0.033	0.016	0.015	0.033	0.031	0.040	0.025	0.024	0.016	0.015	0.022	0.014	0.012

Tabel 10. Standar Deviasi dari LR, AB, DT dengan *Cox* dan TD-Cox Retensi Mahasiswa di UIB pada 2017 (Exp1) untuk Setiap Semester.

Semester	Accuracy					F-Measure					AUC				
	LR	AB	DT	Cox	TD-Cox	LR	AB	DT	Cox	TD-Cox	LR	AB	DT	Cox	TD-Cox
1	0.019	0.017	0.025	0.018	0.018	0.025	0.027	0.030	0.024	0.024	0.016	0.012	0.017	0.014	0.014
2	0.018	0.019	0.025	0.019	0.020	0.029	0.026	0.029	0.021	0.023	0.015	0.013	0.019	0.015	0.016
3	0.019	0.022	0.024	0.015	0.016	0.027	0.025	0.032	0.023	0.021	0.014	0.013	0.018	0.013	0.012
4	0.021	0.018	0.023	0.017	0.014	0.026	0.024	0.031	0.023	0.024	0.014	0.016	0.017	0.011	0.014
5	0.017	0.019	0.021	0.017	0.016	0.023	0.028	0.031	0.025	0.022	0.012	0.014	0.018	0.012	0.012
6	0.018	0.016	0.020	0.014	0.013	0.024	0.027	0.029	0.022	0.020	0.014	0.015	0.017	0.010	0.009
7	0.017	0.015	0.022	0.016	0.015	0.023	0.026	0.030	0.024	0.023	0.017	0.014	0.015	0.012	0.011
8	0.018	0.017	0.021	0.015	0.014	0.025	0.024	0.032	0.026	0.021	0.015	0.016	0.017	0.013	0.010

Tabel 11. Standar Deviasi dari LR, AB, DT dengan *Cox* dan TD-Cox Retensi Mahasiswa di UIB dari 2013-2016 (Exp 2) untuk Setiap Semester dengan 10-fold Cross Validation.

Semester	Accuracy					F-Measure					AUC				
	LR	AB	DT	Cox	TD-Cox	LR	AB	DT	Cox	TD-Cox	LR	AB	DT	Cox	TD-Cox
1	0.023	0.019	0.035	0.015	0.015	0.031	0.028	0.049	0.029	0.029	0.015	0.013	0.026	0.013	0.013
2	0.025	0.026	0.037	0.019	0.018	0.033	0.029	0.051	0.031	0.030	0.018	0.015	0.023	0.013	0.011
3	0.021	0.021	0.036	0.019	0.017	0.036	0.033	0.043	0.029	0.029	0.016	0.014	0.020	0.018	0.014
4	0.023	0.027	0.034	0.018	0.017	0.033	0.035	0.048	0.028	0.029	0.018	0.016	0.023	0.016	0.013
5	0.027	0.030	0.039	0.021	0.019	0.035	0.032	0.041	0.027	0.027	0.016	0.017	0.018	0.015	0.013
6	0.031	0.025	0.035	0.018	0.017	0.033	0.035	0.046	0.026	0.025	0.015	0.015	0.022	0.013	0.010
7	0.029	0.027	0.033	0.020	0.019	0.031	0.033	0.044	0.025	0.026	0.017	0.014	0.020	0.016	0.011
8	0.030	0.025	0.032	0.018	0.018	0.030	0.029	0.043	0.022	0.025	0.014	0.015	0.021	0.014	0.013

Tabel 12. Standar Deviasi dari LR, AB, DT dengan *Cox* dan TD-Cox Retensi Mahasiswa di UIB pada 2017 (Exp2) untuk Setiap Semester.

Semester	Accuracy					F-Measure					AUC				
	LR	AB	DT	Cox	TD-Cox	LR	AB	DT	Cox	TD-Cox	LR	AB	DT	Cox	TD-Cox
1	0.019	0.017	0.025	0.018	0.018	0.025	0.027	0.030	0.024	0.024	0.016	0.012	0.017	0.014	0.014
2	0.018	0.019	0.022	0.019	0.020	0.029	0.026	0.029	0.021	0.023	0.015	0.013	0.019	0.017	0.016
3	0.022	0.024	0.026	0.018	0.017	0.029	0.023	0.035	0.023	0.021	0.015	0.015	0.018	0.014	0.013
4	0.023	0.019	0.024	0.020	0.016	0.031	0.025	0.035	0.025	0.024	0.018	0.018	0.019	0.015	0.016
5	0.019	0.018	0.023	0.018	0.017	0.025	0.029	0.031	0.027	0.022	0.015	0.016	0.021	0.013	0.012
6	0.017	0.019	0.021	0.015	0.014	0.027	0.028	0.029	0.029	0.021	0.013	0.017	0.020	0.014	0.015
7	0.020	0.020	0.019	0.017	0.015	0.028	0.026	0.030	0.025	0.025	0.014	0.018	0.019	0.011	0.014
8	0.021	0.018	0.023	0.016	0.014	0.026	0.025	0.029	0.024	0.026	0.017	0.015	0.020	0.012	0.015

Tabel 13. Hasil dari Metode Regresi Linier, SVR dan Cox pada Prediksi Semester *Dropout* di UIB dari 2013-2016 untuk Setiap Semester dengan 10-fold Cross Section Validation dan Retensi Data Mahasiswa 2017 dengan Kedua Experimen (Exp 1 dan Exp 2)

	Model	10-fold Cross Validation			Data Tes (Tahun 2017)		
		MAE	UPER	OPER	MAE	UPER	OPER
Exp 1	Regression	1.76	0.632	0.368	1.83	0.612	0.388
	SVR	1.89	0.582	0.418	1.91	0.576	0.424
	Cox	1.09	0.697	0.303	1.10	0.676	0.324
Exp 2	Regression	1.92	0.603	0.397	1.96	0.601	0.399
	SVR	1.97	0.576	0.424	2.02	0.571	0.429
	Cox	1.15	0.675	0.325	1.17	0.654	0.346

SIMPULAN DAN SARAN

List atribut yang digunakan dalam penelitian ini yang paling berpengaruh terhadap *Student Dropout* adalah atribut Seleksi Masuk Universitas dan Setelah Masuk Universitas.

Dilakukan tiga klasifikasi model yaitu *Accuracy*, *F-Measure* dan *AUC*. Dari klasifikasi tersebut dilakukan perbandingan metode yaitu *Logistic Regression* (LR), *Adaptive Boosting* (AB), *Decision Tree* (DT), *Cox* dan *Time-Dependent Cox* (TD-Cox). hasilnya diperoleh bahwa metode yang paling bagus adalah *TD-Cox*. Jika dibandingkan antara *Cox* dan *TD-Cox* nilai *performance*-nya sama untuk semester 1 baik untuk Experimen 1 maupun Experimen 2. Secara keseluruhan, metode *TD-Cox* memberikan *performance* yang tinggi untuk *AUC* diikuti *F-Measure* dan *Accuracy*.

Sedangkan untuk nilai standar deviasi dari tiga klasifikasi model juga memberikan kesimpulan yang sama, yaitu metode *TD-Cox* memberikan nilai yang sama dengan metode *Cox* untuk Semester 1 baik untuk Experimen 1 dan Experimen 2 dan yang tertinggi ada dalam klasifikasi *AUC* jika dibandingkan *F-Measure* dan *Accuracy*.

Untuk estimasi *Time to Dropout* digunakan metode *MAE* dengan evaluasi model menggunakan *UPER* dan *OPER*. Dengan adanya data yang tersensor, *Cox* memberikan nilai *UPER* tinggi dibandingkan metode *Regression* dan *SVR*. Hal ini mengindikasikan bahwa error secara keseluruhan berasal dari *underestimating the semester of dropout*.

Berdasarkan hasil penelitian ini dapat penulis menyarankan kepada para pimpinan Program Studi dan Dekan untuk membangun hubungan akademik antara Tenaga Pendidik dengan Mahasiswa terutama mahasiswa semester 1 mengingat sebanyak 8.7% *student dropout* ada disemester ini. Penelitian ini dapat ditindaklanjuti dengan mengembangkan atribut lainnya, sehingga lebih komprehensif dengan metode estimasi lainnya. Memprediksi mahasiswa yang akan *dropout* dari kuliah sangatlah penting dan merupakan tugas yang penuh tantangan bagi institusi pendidikan termasuk UIB. Dengan menjalankan berbagai strategi yang baik akan menurunkan resiko *student dropout*.

UCAPAN TERIMAKASIH

Peneliti menyampaikan rasa terima kasih kepada Kemenristekdikti yang telah memberikan kesempatan dalam Kompetisi Hibah Penelitian. Penelitian ini merupakan bagian dari Penelitian Dosen Pemula Kemenristekdikti Pendanaan Tahun 2017. Selain itu, terima kasih penulis sampaikan kepada LPPM Universitas Internasional Batam yang telah memfasilitasi dalam penelitian ini dengan Kontrak Penelitian Dosen Pemula Nomor : 008/LPPM1.1.1/DP/2017.

DAFTAR PUSTAKA

- [1]A. Nandeshwar, T. Menzies & A. Nelson. 2011. Learning patterns of university student retention. *Expert System with Applications*. 38(12): 14984-14996.
- [2]Astutik, Yayuk S. 2017. Implementasi Cox Proportional Hazard Model Parametrik untuk Analisis Survival Mahasiswa. Laporan Penelitian Hibah Dosen Pemula 2017.
- [3]B. K. Bhardwaj & S. Pal. 2011. A prediction for performance improvement using classification. *International Journal of Computer Science and Information Security*, 9(4): 136-140.
- [4]C. H. Yu, S. DiGangi, A. Jannasch-Pennel & C. Kaprolet. 2010. A data mining approach for identifying predictors of student retention from sophomore to junior year. *Journal of Data Science*, 8:307-325.
- [5]D. Delen. 2011. Predicting student attrition with data mining methods. *Journal of College Student Retention: Research, Theory & Practice*. 13 (1): 17-35.
- [6]J. G. Glynn, P.L. Sauer & T. E. Miller. 2011. A logistic regression model for the enhancement of student retention: The identification of at-risk freshmen. *International Business & Economics Research Journal (IBER)*, 1(8).
- [7]Lawless, J.F. 1982. Statistical Models and Methods for Lifetime Data. *John Wiley & Sons, Inc. USA*.
- [8]M. J. Fard, S. Chawla & C. K. Reddy. 2016. Early-stage event prediction for longitudinal data. In Pacific-Asia

- Conference on Knowledge Discovery and Data Mining, pages 139-151. Springer.
- [9]M. Quadri & N. Kalyankar. 2010. Dropout feature of student data for academic performance using decision tree techniques. *Global Journal of Computer Science and Technology*. 10(2): 1-5.
- [10]Paura, Liga & Arhipova, Iriana. 2014. Cause Analysis of Student Dropout Rate in Higher Education Study Program. *ELSEVIER Journal*, Volume 109, Pages 1282-1286.
- [11]R. Alkhasawneh. 2011. Developing a hybrid model to predict student first year retention and academic success in STEM disciplines using neural network. PhD Thesis, Virginia Commonwealth University.