

Process Mining Akademik Sekolah Menggunakan RapidMiner

Zaki Muhammad¹, Rifqi Rahmadhani², Hafid Rizqifaluthi³, Muhammad Ainul Yaqin⁴

Abstrak— Proses mining adalah sebuah ilmu yang dikembangkan dari data mining. Proses mining mengolah data dalam bentuk event logs yang merupakan representasi dari proses bisnis. Jadi jika sebuah perusahaan atau organisasi sudah memiliki sistem informasi yang menyimpan logs secara otomatis, menerapkan proses mining akan mudah dilakukan. Event logs kemudian dipelajari, dimodelkan untuk menemukan ‘proses model’ yang sesuai dengan kejadian-kejadian yang terekam dalam events log..

Kata Kunci : Data Mining, logs, proses bisnis

I. PENDAHULUAN

Process mining adalah suatu proses pengamatan yang dilakukan terhadap proses bisnis yang telah terkomputerisasi. Proses mining ini erat kaitannya dengan BPM (Business Process Management), dimana BPM itu sendiri bertujuan untuk mendapatkan model bisnis dengan cara mengamati perilaku proses bisnis dalam suatu organisasi. Dengan cara ini diharapkan akan ditemukan struktur proses baru yang sebelumnya tidak disadari sedang terjadi.

Berbagai manfaat bisa didapat dengan adanya Process Mining, seperti untuk mengetahui bagaimanakah proses dalam akademik sekolah, apakah proses yang berjalan sudah sesuai dengan model yang dirancang, maupun mengetahui di tahapan manakah terjadi perlambatan suatu proses. Selain itu Process Mining juga mampu melakukan prediksi atas jumlah keterlambatan yang mungkin timbul serta membuat rancangan model seperti apa yang lebih tepat guna

Manuscript received July 22, 2018. This work was supported in part by Informatics Engineering Department of Maulana Malik Ibrahim Islamic State University.

Zaki Muhammad is with the Informatic Engineering Departement of Maulana Malik Ibrahim Islamic State University , Malang, Indonesia (email zakimuhammad20@gmail.com)

Rifqi Rahmadhani is with the Informatic Engineering Departement of Maulana Malik Ibrahim Islamic State University , Malang, Indonesia (email rifqird98@gmail.com)

Hafid Rizqifaluthi is with the Informatic Engineering Departement of Maulana Malik Ibrahim Islamic State University , Malang, Indonesia (email hafidparaiso@gmail.com)

Muhammad Ainul Yaqin is with the Informatic Engineering Departement of Maulana Malik Ibrahim Islamic State University , Malang, Indonesia (email yaqinov@ti.uin-malang.ac.id)

menyelesaikan permasalahan. Penggunaan event log sebagai sumber data dari teknik Process Mining dirasa cukup tepat karena umumnya log sebuah sistem informasi berisi data dari berbagai kasus yang dieksekusi organisasi. Data yang dicatat umumnya berupa waktu mulai dan selesainya pekerjaan di suatu bagian, siapa saja pelakunya, dan lain sebagainya.

II. TINJAUAN PUSTAKA

A. Proses Mining

Proses Mining merupakan suatu teknologi yang relatif masih baru dalam kaitannya dengan BPM (Business Process Management) [3]. BPM sendiri bertujuan untuk mendapatkan model dengan cara mengamati perilaku proses bisnis di suatu organisasi. Pada proses mining, pengamatan dilakukan terhadap proses bisnis yang telah terkomputerisasi. Dengan ini diharapkan akan ditemukan struktur proses baru yang sebelumnya tidak disadari sedang terjadi. Berdasarkan siklus yang konsisten serta frekuensi aliran nformasi yang terjadi maka dapat diketahui apakah selama ini proses bisnis yang diterapkan oleh sistem informasi telah sesuai dengan pedoman yang dimiliki oleh organisasi ataukah sebaliknya. Berbagai manfaat bisa didapat dengan adanya proses mining, seperti untuk mengetahui bagaimanakah proses yang sebenarnya terjadi. Mengetahui apakah proses yang berjalan sudah sesuai dengan model yang dirancang sebelumnya. Mengetahui di tahapan manakah terjadi perlambatan proses. Selain itu yang cukup menarik bahwasannya aplikasi ini juga mampu melakukan prediksi atas jumlah keterlambatan yang mungkin timbul serta membuat rancangan model seperti apa yang lebih teoat guna menyelesaikan permasalahan.

Process Mining itu sendiri terdiri dari tiga kegiatan utama di dalamnya berdasarkan ada atau tidaknya model terdahulu, yaitu

a. Discovery

Proses ini adalah untuk menemukan model di mana tidak ada model terdahulu yang ditemukan sebelumnya. Berdasarkan event log maka model proses bisnin akan tergambarkan. Penggambaran tersebut berbeda-beda sesuai dengan algoritma yang digunakan.

b. Confermance Analysis

Kegiatan ini dilakukan bila sudah didapatkan model proses bisnis dari proses discovery.

Model proses bisnis ini kemudian dibandingkan dengan event log untuk dianalisis kesesuaian model proses dengan data nyata pada event log.

c. Extention

Pada kegiatan ini model yang terbentuk akan dilakukan pembenahan, yaitu menemukan bottlenecej yang terdapat pada model proses bisnis yang erbentuk. Langkah ini dilakukan bertujuan untuk memperkaya dan membenahi model yang terbentuk.

B. Event Log

Event log merupakan proses pencatatan histori berupa transaksi atau audit trail pada suatu tool sistem informasi. Setiap sistem informasi pasti memiliki event log sebagai bukti dari ransaksi yang sedang berlangsung. Contohnya saja pencatatan event log yang ada pada ERP. Event log berisi informasi tentang kegiatan berupa case atau task tertentu.

Case sendiri disebut dengan “process instance” merupakan suatu kejadian yang sedang berlangsung [5]. Sedangkan task adalah aktivitas di dalam trace, bisa berupa tahapan aktivitas. Jadi dalam trace bisa terdapat banyak task [1]. Event log terditri dari beberapa atribut, di antaranya adalah Case ID, taks, event, user (originator), dan waktu (timestamp).

C. Repidminer Tool

RapidMiner adalah platform perangkat lunak data ilmu pengetahuan yang dikembangkan oleh perusahaan dengan nama yang sama, yang menyediakan lingkungan terpadu untuk pembelajaran mesin (machine learning), pembelajaran mendalam (deep learning), penambangan teks (text mining), dan analisis prediktif (predictive analytics). Aplikasi ini digunakan untuk aplikasi bisnis dan komersial serta untuk penelitian, pendidikan, pelatihan, pembuatan prototype dengan cepat, dan pengembangan aplikasi serta mendukung semua langkah proses pembelajaran mesin termasuk persiapan data, visualisasi hasil, validasi dan pengoptimalan. RapidMiner dikembangkan dengan model open core.

D. Algoritma yang digunakan

A. Alpha Miner

Algoritma alpha memindai event log untuk pola tertentu. Sebagai contoh jika aktivitas **a** diikuti oleh **b** tetapi **b** tidak pernah diikuti oleh **a**, maka diasumsikan bahwa ada ketergantungan kausal antara **a** dan **b**. Untuk merefleksikan dependensi ini, petriNet harus memiliki tempat untuk menghubungkan a ke b.

	<i>a</i>	<i>b</i>	<i>c</i>	<i>d</i>	<i>e</i>
<i>a</i>	# _{L1}	→ _{L1}	→ _{L1}	# _{L1}	→ _{L1}
<i>b</i>	← _{L1}	# _{L1}	_{L1}	→ _{L1}	# _{L1}
<i>c</i>	← _{L1}	_{L1}	# _{L1}	→ _{L1}	# _{L1}
<i>d</i>	# _{L1}	← _{L1}	← _{L1}	# _{L1}	← _{L1}
<i>e</i>	← _{L1}	# _{L1}	# _{L1}	→ _{L1}	# _{L1}

Table 2.1 Footprint of L₁: a # L₁ a, a → L₁ b, a → L₁ c, etc

- a >L b jika dan hanya jika jejak σ = {t1,t2,t3,...,tn} dan i ∈ {1,...,n - 1} seperti σ ∈ L dan ti = a and ti+1 = b
- a →L b jika dan hanya jika a >L b and b >L a
- a #L b jika dan hanya jika a >L b and b >L a
- a L b jika dan hanya jika a >L b and b >L a

Pertimbangan misalnya, L1 = [{a,b,c,d}3,{a,c,b,d}2,{a,e,d}] lagi. Untuk event log ini, log-based yang mengikuti relasi yang diminta dapat ditemukan.

$$>L1 = \{(a,b),(a,c),(a,e),(b,c),(c,b),(b,d),(c,d),(e,d)\}$$

$$\rightarrow L1 = \{(a,b),(a,c),(a,e),(b,d),(c,d),(e,d)\}$$

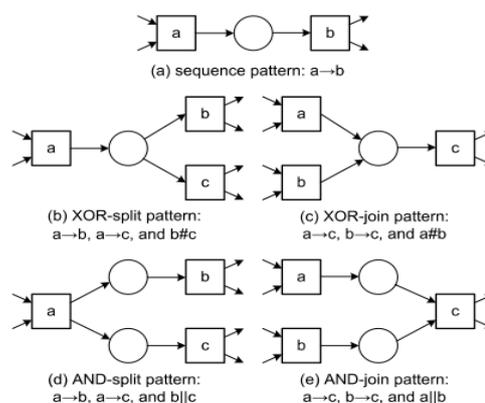
$$\#L1 = \{(a,a),(a,d),(b,b),(b,e),(c,c),(c,e),(d,a),(d,d),(e,b),(e,c),(e,e)\}$$

$$\|L1 = \{(b,c),(c,b)\}$$

Hubungan >L1 memuat semua pasangan akitifitas pada relasi “directly follows”. c > L1 d karena d secara langsung mengikuti c pada trace (a,b,c,d). Bagaimanapun, d tidak >L1 c karena c tidak pernah secara langsung mengikuti d pada jejak manapun didalam log. →L1 memuat seluruh pasangan aktivitas pada hubungan “kausalitas”, misalnya, c →L1 d karena terkadang d secara langsung c dan tidak pernah sebaliknya (c >L1 d and d tidak >L1 c). b || L1 c karena b > L1 c dan c > L1 b yaitu, terkadang c mengikuti b dan terkadang sebaliknya b#L1 e karena b tidak > L1 e dan e tidak >1 b.

	<i>a</i>	<i>b</i>	<i>c</i>	<i>d</i>	<i>e</i>	<i>f</i>
<i>a</i>	#	→	→	#	#	#
<i>b</i>	←	#		→	→	←
<i>c</i>	←		#	→	→	←
<i>d</i>	#	←	←	#	#	#
<i>e</i>	#	←	←	#	#	→
<i>f</i>	#	→	→	#	←	#

Table 2.2



Gambar 2.3

Relasi order log-based dapat digunakan untuk menemukan pola dalam model proses yang sesuai

seperti yang diilustrasikan pada Gambar. 5.4. Jika a dan b berurutan, log akan menunjukkan $a \rightarrow L b$. Jika setelah ada pilihan antara b dan c, log akan menunjukkan $a \rightarrow L b$, $a \rightarrow L c$, dan $b \#L c$ karena a dapat diikuti oleh b dan c, tetapi b tidak akan diikuti oleh c dan sebaliknya. Pendapat logis dari pola XOR-split ini adalah pola XOR-join seperti ditunjukkan pada Gambar. 5.4 (b -c). Jika $a \rightarrow L c$, $b \rightarrow L c$, dan $a \#L b$, maka ini menunjukkan bahwa setelah terjadinya a atau b, c harus terjadi. Gambar 5.4 (d-e) menunjukkan apa yang disebut pola AND-split dan AND-join. Jika $a \rightarrow L b$, $a \rightarrow L c$, dan $b L c$, maka tampak bahwa setelah b dan c dapat dieksekusi secara paralel (pola AND-split). Jika $a \rightarrow L c$, $b \rightarrow L c$, dan $L b$, maka log menunjukkan bahwa c perlu mensinkronisasi a dan b (AND-join pattern).

B. Fuzzy Mining

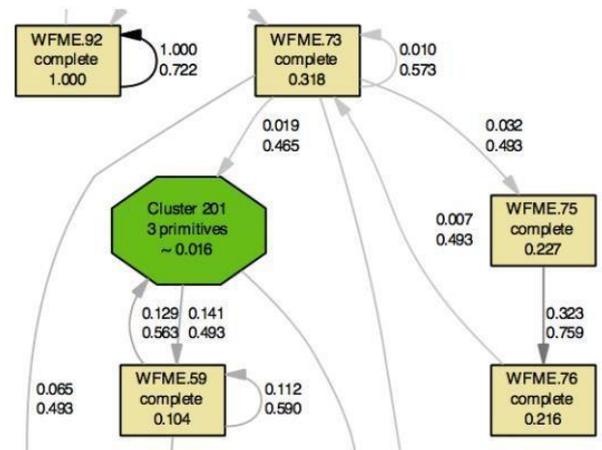
Fuzzy mining merupakan salah satu teknik dalam melakukan *process mining* yang ditujukan pada *real-life log*. Proses yang terjadi pada *real-life log* kurang terstruktur dimana pendekatan *process mining* seperti *alpha* dan *heuristic* memiliki masalah terhadap data yang kurang terstruktur tersebut. Model proses yang dihasilkan oleh pendekatan *alpha* dan *heuristic* sering “*spaghettilike*”, menampilkan semua detail tanpa membedakan proses mana yang penting dan mana tidak penting. Oleh karena itu, pendekatan *fuzzy mining* diterapkan untuk mengatasi model proses yang “*spaghetti-like*” tersebut. Untuk melakukannya, konsep peta jalan digunakan sebagai perumpamaannya untuk memvisualisasikan model yang dihasilkan [1].



Gambar 2.4 Konsep Peta Jalan

Dalam konsep peta jalan Gambar 2.4, dapat dilihat bahwa komponen-komponen yang memiliki peran kurang penting dalam suatu peta jalan tidak akan ditampilkan seperti jalan-jalan kecil. Beberapa bagian kota kecil yang masih dalam satu bagian kota besar akan digabungkan. Pembuatan peta jalan tersebut dan tingkat kerinciannya hanya difokuskan untuk tujuan penggunaannya saja. Konsep tersebut diterapkan pada pendekatan *fuzzy mining* dalam melakukan *process mining*. Untuk melakukan hal tersebut, diperlukan kriteria keputusan yang tepat sebagai dasar penyederhanaan dan visualisasi suatu model proses. Terdapat dua *metrics* dasar yang dapat mendukung keputusan tersebut yaitu *significance* dan *correlation*. *Significance* dapat ditentukan untuk kegiatan (*node*) dan relasi (*edge*) dalam suatu proses, dengan mengukur nilai *relative importance*-nya. Dengan demikian, hal ini dapat

menentukan tingkat kepentingan yang dimiliki oleh suatu kegiatan. Salah satu contoh mengukur *significance* adalah dengan frekuensi, yaitu kegiatan atau relasi yang diamati lebih sering dianggap signifikan. Disisi lain, *correlation* hanya relevan untuk relasi yang lebih diutamakan daripada kegiatannya. *Correlation* mengukur seberapa erat hubungan antar dua kegiatan berikut dengan satu sama lainnya [1].



Gambar 2.5 Kutipan Model Proses yang Sederhana

III. METODE PENELITIAN

A. Event Log Akademik Sekolah

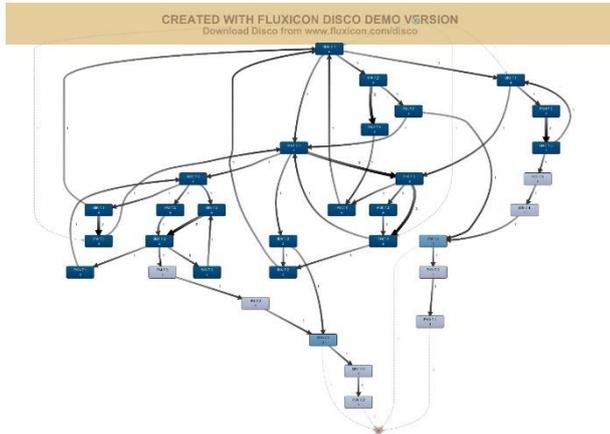
Untuk melakukan proses bisnis Akademik Sekolah tentunya harus membuat Event log untuk proses pencatatan histori berupa transaksi atau audit trail pada suatu tool sistem informasi. Event log Akademik Sekolah yang dibuat menggunakan 100 baris, 4 case dan setiap case mempunyai 25 event.

Case ID	Event ID	Timestamp	Activity	Resource	Ruang
1	1	16/07/2018	SBD 7.1	Guru 10	101
1	2	16/07/2018	MAT 7.1	Guru 5	101
1	3	16/07/2018	PJK 7.1	Guru 7	101
1	4	16/07/2018	PAI 7.1	Guru 6	101
1	5	17/07/2018	MAT 7.2	Guru 5	101
1	6	17/07/2018	BIN 7.1	Guru 2	101
1	7	17/07/2018	IPA 7.1	Guru 3	101
1	8	17/07/2018	SBD 7.2	Guru 10	101
1	9	18/07/2018	PKN 7.1	Guru 8	101
1	10	18/07/2018	BIN 7.2	Guru 2	101
1	11	18/07/2018	MAT 7.3	Guru 5	101
1	12	18/07/2018	BIG 7.1	Guru 1	101
1	13	19/07/2018	BIG 7.2	Guru 1	101
1	14	19/07/2018	IPS 7.1	Guru 4	101
1	15	19/07/2018	PKN 7.2	Guru 8	101
1	16	19/07/2018	BIN 7.3	Guru 2	101
1	17	20/07/2018	PJK 7.2	Guru 7	101
1	18	20/07/2018	IPS 7.2	Guru 4	101
1	19	20/07/2018	IPA 7.2	Guru 3	101
1	20	20/07/2018	PKY 7.1	Guru 9	101
1	21	20/07/2018	PAI 7.2	Guru 6	101
1	22	21/07/2018	IPA 7.3	Guru 3	101
1	23	23/07/2018	SBD 7.3	Guru 10	101
1	24	23/07/2018	MAT 7.4	Guru 5	101

Gambar 3.1 Event Log Akademik Sekolah

B. Estensi Data

Untuk melakukan proses discoveri pada RapidMiner perlu dilakukan ekstensi data. Ekstensi data yang dilakukan ialah melakukan ekstensi data dari format .csv ke format .xes. Format inilah yang nantinya digunakan untuk dilakukan pemodelan proses discovery menggunakan aplikasi proses mining yaitu RapidMiner. Adapun dilakukan dengan menggunakan aplikasi ProM Import atau Disco. Aplikasi tersebut yang akan mengubah dari format .csv menjadi format .xes.

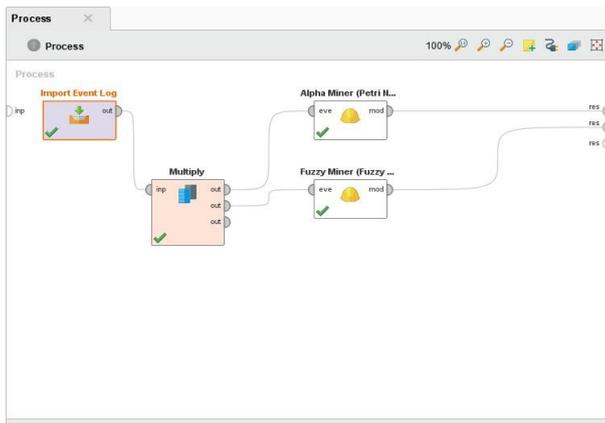


Gambar 3.2 Estensi Data Event Log dari format .csv menjadi format .xes.

C. Pemodelan menggunakan Algoritma α dan Fuzzy Miner

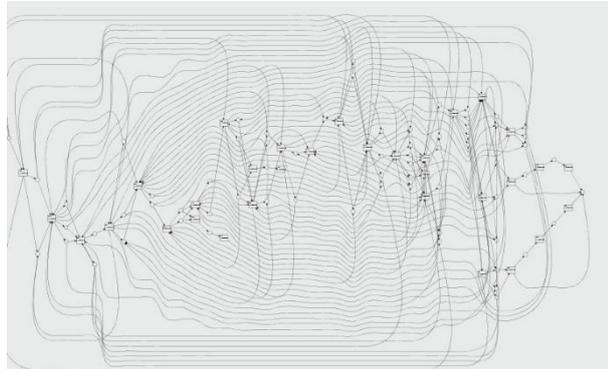
Setelah mendapatkan data maka selanjutnya adalah dilakukan pemodelan berdasarkan data yang telah didapatkan sebelumnya. Pemodelan ini dilakukan dengan menggunakan aplikasi Rapid Miner pada system operasi Windows. Sedangkan untuk algoritma yang digunakan menggunakan algoritma α , dan juga Fuzzy Miner. Kedua algoritma ini dipilih karena memiliki kesesuaian dengan data yang didapat. Dari pemodelan inilah yang nantinya akan dilihat bagaimana kedua algoritma ini bekerja dalam membentuk sebuah model proses bisnis. Dan juga akan dilihat pula ketepatan dari dua algoritma yang ada dalam memodelkan sebuah proses bisnis berdasarkan data log yang ada.

Pemodelan dilakukan di aplikasi RapidMiner dengan cara import data Event log yang berekstensi .xes.

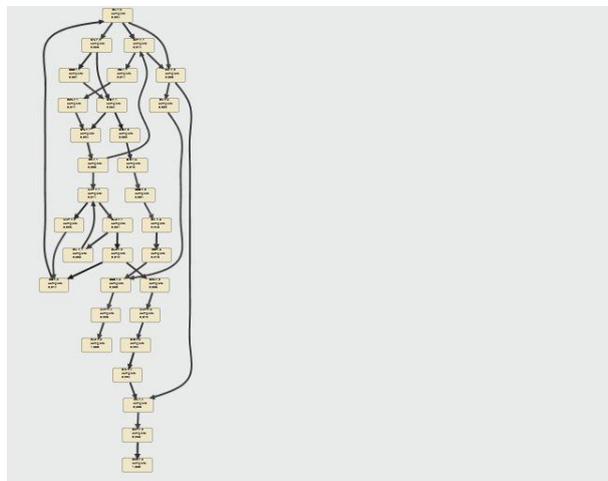


Gambar 3.3 Pemodelan Proses Bisnis Dengan Software RapidMiner

Setelah data masuk maka mulai membuat pemodelan pada data menggunakan 2 algoritma yang berbeda :



Gambar 3.4 Pemodelan menggunakan algoritma α (alpha).



Gambar 4.3 Pemodelan menggunakan algoritma Fuzzy Miner

IV. PEMBAHASAN

Setelah dicoba proses mining akademik sekolah dengan RapidMiner, dengan menggunakan 2 algoritma untuk membandingkan, yaitu dengan algoritma Alpha Miner dan algoritma Fuzzy Miner. Didapatkan hasil bahwa hasil visualisasi proses mining dari algoritma fuzzy miner lebih sederhana dan mudah dibaca dibandingkan dengan menggunakan algoritma alpha miner.

Algoritma fuzzy miner ini didasari pada algoritma fuzzy adaptive graph simplification yang dipercaya dapat lebih menyederhanakan model proses bisnis yang dihasilkan sebagai output sehingga mudah dimengerti dan mudah pula untuk dianalisa. Jika dibandingkan dengan algoritma alpha miner, proses mining menggunakan algoritma selalu menghasilkan model proses bisnis yang sangat kompleks dan tidak terstruktur (Spaghety like case). Makadari itu hal inilah yang menjadi fokus dari dilakukannya Process mining dengan menggunakan Fuzyy Mining yaitu untuk menampilkan model proses bisnis yang sederhana namun memiliki informasi yang sesuai dengan kebutuhan.

VI. KESIMPULAN

Pada pengujian model yang terbentuk maka dapat diketahui bahwa algoritma yang digunakan berperan dalam menggambarkan ketetapan penggambaran proses bisnis ERP. Penggambaran model dari Algoritma Fuzzy Miner tergambar paling baik daripada algoritma α karena algoritma Fuzzy Miner dapat menangani Sepagethi-like, sehingga penggambarannya lebih tepat dari algoritma α .

REFERENCES

- [1] Ritchi, H. 2009. *Identifikasi Pengendalian Aplikasi dalam Analisis Proses Bisnis*. Bandung: Pustaka UNPAD.
- [2] Weber, P. 2010. *Research Progress Report 3 – Thesis Proposal Working Title of Thesis: Machine Learning in Process Mining*. School of Computer Science University of Brimingham.
- [3] Weijers, A., van der. 2008. *Process Mining: Beyond Business Intelligence*. <http://www.processmining.org>
- [4] perduti, A. B. 2010. *Automatic Determination of Parameters' Value for Heuristics Miner*. Italy: Departement of Pure and Applied Mathematics University of Padua.
- [5] Weber, P. 2009. *A Framework for The Comparison of Process Mining Algorithms*. School of Computer Science University of Brimingham
- [6] Marlon Dumas, Marcello La Rosa, Jan Mendling, Hajo A. Reijers, 2013. *Fundamentals of Business Process Management*. New York: Springer.
- [7]. Wil M. P. Van der Aalst. 2011. *Proces Mining, Discovery, Conformance and Enhancement of Business Processes*. New York: Springer.
- [8]. A W Prestya. M A Yaqin. S Zaman. 2018. *Common oses process Extraction Pada Scalable Model Proses Bisnis*. STMK ATMA LUHUR.