

Πανεπιστήμιο Δυτικής Αττικής

Τμήμα Μηχανικών Πληροφορικής και  
Υπολογιστών



Διπλωματική Εργασία

Αξιολόγηση διαδικασιών  
Ευρωπαϊκών χρηματοδοτήσεων με  
χρήση αλγορίθμων εξόρυξης  
διαδικασιών

Εισηγητής:

Αλέξανδρος Μπουσδέκης  
ΠΑΝΕΠΙΣΤΗΜΙΑΚΟΣ  
ΥΠΟΤΡΟΦΟΣ ΕΣΠΑ

Υποψήφιος:

Γκούσαρης Κωνσταντίνος  
Α.Μ 7111171073

---

Ακαδημαϊκό Έτος 2022/2023



## ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

Αξιολόγηση διαδικασιών Ευρωπαϊκών χρηματοδοτήσεων  
με χρήση αλγορίθμων εξόρυξης διαδικασιών

Γκουσαρης Κωνσταντίνος - 711171073 - cs171073@uniwa.gr

Εισηγητής: Μπουσδέκης Αλέξανδρος, ΠΑΝΕΠΙΣΤΗΜΙΑΚΟΣ  
ΥΠΟΤΡΟΦΟΣ ΕΣΠΑ

Εξεταστική Επιτροπή: Μαστοροκώστας Πάρις, ΚΑΘΗΓΗΤΗΣ,  
Μιαούλης Γεώργιος, ΟΜΟΤΙΜΟΣ ΚΑΘΗΓΗΤΗΣ,  
Μπουσδέκης Αλέξανδρος, ΠΑΝΕΠΙΣΤΗΜΙΑΚΟΣ  
ΥΠΟΤΡΟΦΟΣ ΕΣΠΑ

Ημερομηνία εξέτασης: 24/07/2023

Μαστοροκώστας  
Πάρις,  
ΚΑΘΗΓΗΤΗΣ

Μιαούλης  
Γεώργιος  
ΟΜΟΤΙΜΟΣ  
ΚΑΘΗΓΗΤΗΣ

Μπουσδέκης  
Αλέξανδρος  
ΠΑΝΕΠΙΣΤΗΜΙΑΚΟΣ  
ΥΠΟΤΡΟΦΟΣ  
ΕΣΠΑ

---

Υπογραφή

---

Υπογραφή

---

Υπογραφή

Αθήνα, Ιούλιος 2023



## Δ'ΗΛΩΣΗ ΣΥΓΓΡΑΦ'ΕΑ ΔΙΠΛΩΜΑΤΙΚ'ΗΣ ΕΡΓΑΣΪΑΣ

Ο κάτωθι υπογεγραμμένος Γκούσαρης Κωνσταντίνος του Ηλία, με αριθμό μητρώου 711171073 φοιτητής του Πανεπιστημίου Δυτικής Αττικής της Σχολής Μηχανικών του Τμήματος Μηχανικών Πληροφορικής και Υπολογιστών, δηλώνω υπεύθυνα ότι:

«Είμαι συγγραφέας αυτής της πτυχιακής/διπλωματικής εργασίας και ότι κάθε βοήθεια την οποία είχα για την προετοιμασία της είναι πλήρως αναγνωρισμένη και αναφέρεται στην εργασία. Επίσης, οι όποιες πηγές από τις οποίες έκανα χρήση δεδομένων, ιδεών ή λέξεων, είτε ακριβώς είτε παραφρασμένες, αναφέρονται στο σύνολό τους, με πλήρη αναφορά στους συγγραφείς, τον εκδοτικό οίκο ή το περιοδικό, συμπεριλαμβανομένων και των πηγών που ενδεχομένως χρησιμοποιήθηκαν από το διαδίκτυο. Επίσης, βεβαιώνω ότι αυτή η εργασία έχει συγγραφεί από μένα αποκλειστικά και αποτελεί προϊόν πνευματικής ιδιοκτησίας τόσο δικής μου, όσο και του Ιδρύματος. Παράβαση της ανωτέρω ακαδημαϊκής μου ευθύνης αποτελεί ουσιώδη λόγο για την ανάκληση του πτυχίου μου».

Ο Δηλών: Κωνσταντίνος Γκούσαρης  
Γκούσαρης Κωνσταντίνος



*If you can't describe what you are doing as a process, you don't know what you're doing.*

- W.Edwards Deming





## Ευχαριστίες

Στην οικογένεια μου και στην σύζυγο μου που με στήριξαν σε αυτό το ταξίδι επανεκπαίδευσης, στους φίλους και συμφοιτητές που βοήθησαν να γίνει αυτό εφικτό, στους καθηγητές μου γιατί μέσα από την διαδικασία μάθησης μου αλλάζανε τον τρόπο με τον οποίο αντιλαμβάνομαι τον κόσμο και στον καθηγητή μου Αλέξανδρο Μπουσδέκη για την έμπνευση και την βοήθεια ώστε να έρθει σε πέρας το συγκεκριμένο έργο.



## Abstract

Βασική ιδέα της διπλωματικής είναι να παρουσιάσει τις δυνατότητές της ανάλυσης διαδικασιών με την χρήση τεχνικών του τομέα της εξόρυξης διαδικασιών (Process Mining). Η ανάλυση θα περιέχει τρία σκέλη, την εξόρυξη διαδικασιών παρουσιάζοντας τις σημαντικές πτυχές της επιστήμης, την εφαρμογή της εξόρυξης διαδικασιών με την χρήση της βιβλιοθήκης pm4py της γλωσσάς προγραμματισμού Python και την ανάλυση και αξιολόγηση μιας διαδικασίας αίτησης για επιδότηση από δημόσιο φορέα.

Συγκεκριμένα, θα αναλυθεί το θεωρητικό υπόβαθρο του τομέα της Εξόρυξης Διαδικασιών, παρουσιάζοντας τη δομή των αρχείων καταγραφής γεγονότων, τους πιο σημαντικούς αλγόριθμους εξόρυξης διαδικασιών, τις βασικές τεχνικές ανάλυσης, τις διάφορες σημειογραφίες και τα εργαλεία λογισμικού τα οποία πλαισιώνουν ένα έργο εξόρυξης διαδικασιών. Η εργασία τεκμηριώνει την ανάγκη που υπάρχει σήμερα για σαφείς και ευέλικτες διαδικασίες και το πόσο σημαντικό και απαραίτητο είναι για τους ενδιαφερόμενους οργανισμούς να επενδύσουν σε τέτοιου είδους τεχνογνωσία και εργαλεία. Σε τεχνικό επίπεδο με την χρήση των βιβλιοθηκών της Python, κυρίως της pm4py, έχει δημιουργηθεί πηγαίος κώδικας ο οποίος πραγματοποιεί εξόρυξη και ανάλυση διαδικασιών σε παραγωγικά δεδομένα.

Τα δεδομένα στα οποία επικεντρώνεται η εργασία αφορούν την κατανόηση και την ανάλυση της διαδικασίας που διεκπεραιώνει το πρόγραμμα άμεσων πληρωμών του Ευρωπαϊκού Ταμείου Εγγυήσεων και το ποσό που έχει δεσμευθεί για την αγροτική ανάπτυξη. Οι άμεσες πληρωμές αποτελούν ένα κύριο στοιχείο της κοινής γεωργικής πολιτικής (CAP) της ΕΕ, παρέχοντας οικονομική στήριξη στους αγρότες και προωθώντας την ανταγωνιστικότητα, τη βιωσιμότητα και τις πρακτικές αγροτικής καλλιέργειας που σέβονται το περιβάλλον. Τα δεδομένα προέρχονται από πληθώρα πληροφοριακών συστημάτων σε μορφή event logs και αναλύονται χρησιμοποιώντας διάφορα εργαλεία εξόρυξης και ανάλυσης προκειμένου να εξαχθούν και να αξιολογηθούν τα μοντέλα διαδικασίας που προκύπτουν. Σκοπός είναι, όσο είναι δυνατό, να εντοπιστούν αδύναμα σημεία και να βελτιωθούν οι διαδικασίες.

**Λεξεις κλειδιά:** Process Mining, pm4py, εξόρυξη διαδικασιών, εργαλεία εξόρυξης διαδικασιών, Ανακάλυψη διαδικασιών, αλγόριθμοι ανακάλυψης διαδικασιών, έλεγχος συμμόρφωσης, εμπλουτισμός, αρχείο καταγραφής γεγονότων, κοινή αγροτική πολιτική, Python.



## Abstract

The aim of the thesis is to present the capabilities of process analysis when applying techniques from the field of Process Mining. The analysis will consist of three parts: process mining, presenting the important aspects of the science; the application of process mining using the pm4py library of the Python programming language; and the analysis and evaluation of a grant application process from a public entity.

Specifically, the theoretical background of Process Mining will be examined, including the structure of event log files, the most significant process mining algorithms, basic analysis techniques, various notations, and software tools that encompass a process mining project. The paper highlights the need for clear and flexible processes today and emphasizes how important and necessary it is for organizations to invest in such knowledge and tools.

At a technical level, using Python libraries, mainly pm4py, source code has been created that performs process mining and analysis on real data. The focus of the study is on understanding and analyzing the process carried out by the European Guarantee Fund's direct payment program and the amount allocated for rural development. Direct payments are a key element of the EU's Common Agricultural Policy (CAP), providing financial support to farmers and promoting competitiveness, sustainability, and environmentally friendly agricultural practices.

The data used in this study comes from a variety of information systems in the form of event logs, and it is analyzed using various mining and analysis tools to extract and evaluate the resulting process models. The goal is to identify weaknesses and improve the effectiveness of processes .

**Key Words:** Process Mining, pm4py, Process Mining Software, Process Discovery, Process Discovery algorithms, Conformance Checking, Enhancement, Event Log, IEEE XES, CAP, Python.





# Contents

<b>1</b>	<b>Εισαγωγή</b>	<b>5</b>
<b>2</b>	<b>Business Process Management</b>	<b>7</b>
2.1	Γιατί είναι απαραίτητο το BPM . . . . .	8
2.2	Κύκλος Ζωής BPM . . . . .	9
2.3	Συστήματα BPM . . . . .	12
<b>3</b>	<b>Εξόρυξη Διαδικασιών</b>	<b>15</b>
3.1	Event Logs . . . . .	17
3.2	Πρότυπα αρχείων - XES . . . . .	21
3.3	Μοντέλα διαδικασιών . . . . .	23
3.3.1	PetriNet . . . . .	23
3.3.2	BPMN . . . . .	25
3.3.3	Directly-Follows Graph . . . . .	27
3.3.4	Δέντρα διαδικασιών . . . . .	27
3.4	Αλγόριθμοι Ανακάλυψης Διαδικασιών . . . . .	28
3.4.1	Alpha Miner . . . . .	30
3.4.2	Inductive Miner . . . . .	31
3.4.3	Heuristic Miner . . . . .	32
3.5	Εργασίες εξόρυξης διαδικασιών . . . . .	33
3.5.1	Ανακάλυψη της διαδικασίας (Process Discovery). . . . .	34
3.5.2	Έλεγχος συμμόρφωσης (Conformance Checking). . . . .	35
3.5.3	Βελτιστοποίηση (Enhancement). . . . .	38
3.5.4	Ανάλυση απόδοσης (Performance Analysis). . . . .	39
<b>4</b>	<b>Εργαλεία Εξόρυξης Διαδικασιών</b>	<b>41</b>
4.1	Pm4py . . . . .	43
4.2	Disco . . . . .	44
4.3	ProM . . . . .	44
<b>5</b>	<b>Διαδικασία Χρηματοδοτήσεων Κοινής Αγροτικής Πολιτικής</b>	<b>47</b>
5.1	Dataset . . . . .	47
5.2	Προκλήσεις Διαδικασίας . . . . .	49
<b>6</b>	<b>Ερευνητική Μεθοδολογία</b>	<b>51</b>
6.1	Import Event Log . . . . .	51
6.2	Κατανόηση Διαδικασίας μέσω των δεδομένων . . . . .	51
6.3	Process Discovery . . . . .	52
6.4	Conformance Checking . . . . .	52



6.5	Αξιολόγηση Αλγόριθμων Εξόρυξης Διαδικασιών . . . . .	53
6.6	Ανάλυση Διαδικασίας . . . . .	53
<b>7</b>	<b>Τεχνική Υλοποίηση</b>	<b>55</b>
7.1	Λειτουργικές Απαιτήσεις . . . . .	55
7.2	Περιγραφή Συναρτήσεων . . . . .	55
7.2.1	Γενικές Συναρτήσεις . . . . .	56
7.2.2	Εξαγωγή - εισαγωγή δεδομένων . . . . .	57
7.2.3	Στατιστικά . . . . .	57
7.2.4	Φιλτράρισμα . . . . .	60
7.2.5	Ανακάλυψη Διαδικασίας . . . . .	61
7.2.6	Έλεγχος συμμόρφωσης . . . . .	64
7.2.7	Βελτιστοποίηση και ανάλυση . . . . .	66
<b>8</b>	<b>Εφαρμογή σε διαδικασίες χρηματοδότησης της Κοινής Αγροτικής Πολιτικής</b>	<b>69</b>
8.1	Παρουσίαση Δεδομένων . . . . .	69
8.2	Ανακάλυψη Διαδικασίας . . . . .	75
8.3	Έλεγχος Συμμόρφωσης . . . . .	90
8.4	Αξιολόγηση Μοντέλου . . . . .	101
8.5	Εξόρυξη Κοινωνικών Δικτύων . . . . .	102
8.6	Ανάλυση Decision Point Mining . . . . .	109
<b>9</b>	<b>Συμπεράσματα και Μελλοντική Εργασία</b>	<b>111</b>
9.1	Συμπεράσματα . . . . .	111
9.2	Μελλοντική Εργασία . . . . .	113

# Εισαγωγή

Η χρήση των πληροφοριακών συστημάτων και το γεγονός ότι οι οργανισμοί στηρίζουν την ανάπτυξη και την οικονομική τους ευρωστία σε εργαλεία λογισμικού τα οποία κάνουν την εργασία παραγωγική και γρήγορη έχει δημιουργήσει τομείς και επαγγέλματα συνδυάζοντάς στοιχεία από διαφορετικά επιστημονικά πεδία. Ο τομέας της Διαχείρισης Επιχειρηματικών Διαδικασιών (BPM) είναι ένας από αυτούς και κομμάτι του κύκλου εργασιών του είναι η εξόρυξη διαδικασιών (Process Mining). Η συγκεκριμένη τεχνική έχει δικό της, ανεξάρτητο κύκλο εργασιών στο πλαίσιο ενός ευρύτερου έργου BPM και χρησιμοποιείται για να βελτιώσει την συμμόρφωση των πληροφοριακών συστημάτων, των επιχειρησιακών διαδικασιών με την πολιτική του οργανισμού και τους κανόνες των θεσμών στους οποίους λειτουργεί. Η εξόρυξη διαδικασιών αποδεδειγμένα προσθέτει αξία στους οργανισμούς αναλύοντας τις πραγματικές διαδικασίες, ανακαλύπτοντας αντιφάσεις και προτείνοντας νέες ιδέες που μπορούν να βελτιώσουν την υποστήριξη των εργαζόμενων από τα πληροφοριακά συστήματα.

Χαρακτηριστικό παράδειγμα αποτελεί το γεγονός ότι παρά τις προκλήσεις που αντιμετώπιστηκαν όταν παρουσιάστηκε απότομη πτώση της παραγωγικότητάς και των εσόδων των επιχειρήσεων κατά το ξέσπασμά της πανδημίας, οι περισσότερες εταιρίες κατάφεραν να ανακάμψουν γρήγορα. Αυτό οφείλεται, κυρίως, στην ικανότητα την οποία είχαν αναπτύξει επενδύοντας στα αντίστοιχα τμήματα ψηφιοποιώντας τις επιχειρηματικές διαδικασίες. Με το αυτό το δεδομένο σαν αφετηρία οι επιχειρήσεις γίνονται ψηφιακά ευέλικτες και, συνειδητά, υιοθετούν τεχνολογίες και λύσεις λογισμικού για να ενισχύσουν τις επιχειρηματικές τους δραστηριότητες και να επιτύχουν αποτελεσματικότητα στην εκτέλεση των διαδικασιών τους.

Η πληθώρα εργαλείων που χρησιμοποιούνται για την διερεύνηση διαδικασιών, συνδυάζει ανάλυση και εξόρυξη δεδομένων για να βοηθήσει τους οργανισμούς να εξάγουν δεδομένα από τα συστήματά τους. Ένα από τα πολλά εργαλεία που χρησιμοποιούνται σε ακαδημαϊκό αλλά και επαγγελματικό επίπεδο είναι η pm4py. Πρόκειται για μια βιβλιοθήκη της γλώσσας προγραμματισμού Python η οποία μπορεί να εκτελέσει εργασίες ανακάλυψης διαδικασιών. Πλεονεκτήματά της αποτελούν η ευελιξία, η δυνατότητα χρήσης σε μεγαλύτερα project τα οποία θέλουν να αναλύσουν και να αξιολογήσουν τις διαδικασίες και την ροή εκτέλεσης των δραστηριοτήτων. Αντίθετα αρνητικό πρόσημο φέρουν ο χρόνος επεξεργασίας των δεδομένων, ειδικά για πολύπλοκες διαδικασίες και το γεγονός ότι για να γίνει εφαρμογή της πρέπει ο χρήστης να έχει αρκετά καλές γνώσεις προγραμματισμού και ως ένα βαθμό γνώσεις εξόρυξης δεδομένων. Η pm4py παρέχει διάφορες τεχνικές εξόρυξης διαδικασιών και μοντελοποίησης της διαδικασίας σε πολλά επίπεδα αφαίρεσης. Για παράδειγμα, οι χρήστες μπορούν να αποφασίσουν να εξετάσουν την πιο συχνή συμπεριφορά για να αποκτήσουν ένα απλό μοντέλο ("μοντέλο 80%") αλλά ταυτόχρονα μπορούν επίσης να εξετάσουν την πλήρη συμπεριφορά αποκτώντας το "μοντέλο 100%" που καλύπτει όλες τις παρατηρηθείσες περιπτώσεις.

Με παρόμοιο τρόπο, τα επίπεδα αφαίρεσης μπορούν να διαφοροποιούνται και στις υπόλοιπες εργασίες εξόρυξης διαδικασιών που υποστηρίζει, όπως ο έλεγχος συμμόρφωσης, η βελτιστοποίηση κ.α. Στα πλεονεκτήματα είναι ότι μπορεί να αποκαλύψει την ανθρώπινη συμπεριφορά, να ανακαλύψει καθυστερήσεις και να απεικονίσει την εκπληκτική ευελιξία ορισμένων εργαζομένων στην αντιμετώπιση προβλημάτων.

Στο πλαίσιο της εφαρμογής της θεωρίας της εξόρυξης διαδικασιών θα γίνει εξέταση μιας διαδικασίας αίτησης για χρηματοδότηση επιχορήγησης από τον Ευρωπαϊκό Ταμείο Στήριξης με την χρήση της βιβλιοθήκης της pm4py. Οι συγκεκριμένες διαδικασίες παρουσιάζουν ιδιαίτερο ενδιαφέρον γιατί καθορίζονται από αυστηρά πλαίσια και κανόνες οι οποίοι μεταβάλλονται συνεχώς. Επιπλέον, υπάρχει ο ανθρώπινος παράγοντας, ο οποίος σε τέτοιου είδους διαδικασίες παίζει καθοριστικό ρόλο τόσο στην εκτέλεση όσο και στο αποτέλεσμα της. Τέλος, τα πληροφοριακά συστήματα που συμμετέχουν σε τέτοιες διαδικασίες έχουν κυρίως βοηθητικό ρόλο. Για παράδειγμα στην διαδικασία που εξετάζεται, την επικύρωση της ορθότητας ή όχι των στρεμμάτων που έχει δηλώσει ο αγρότης επικυρώνεται μετά από αυτοψία ενός μηχανικού και καταχώρηση της απόφασής του στο πληροφοριακό σύστημα.

# Business Process Management

Η καινοτομία και η πρόοδος που έχουν αναπτυχθεί τα τελευταία χρόνια στους κλάδους της πληροφορικής, της επιστήμης των δεδομένων και των επικοινωνιών, έχουν επιφέρει αλλαγές στους οργανισμούς και αποτελούν κύριους μοχλούς ανάπτυξης των επιχειρήσεων. Επίσης κοινωνικές αλλαγές, καθοριστικά και ταυτόχρονα αναπάντεχα γεγονότα των τελευταίων ετών, όπως η πανδημία, ο πόλεμος στην Ουκρανία κλπ., επηρεάζουν την λειτουργία των επιχειρήσεων, αναγκάζοντας τις να μεταβάλουν και να αλλάζουν συνεχώς με ταχύτητα και ταυτόχρονα τις επιχειρησιακές τους διαδικασίες. Δεν είναι τυχαίο, ότι έρευνες [17],[10] που πραγματοποιήθηκαν στα μέσα του 2020 για τις συνέπειες της πρόσφατης πανδημίας σε οργανισμούς, έδειξε ότι το 54% των επιχειρήσεων αύξησαν τον προϋπολογισμό τους για αγορά ή αναβάθμιση πληροφοριακών συστημάτων αυτοματισμού διαδικασιών, κύριο συστατικό της Διοίκησης Επιχειρηματικών Διαδικασιών. Διαφαίνεται λοιπόν η ανάγκη για τους οργανισμούς να ακολουθούν τις εξελίξεις, να δημιουργούν διαδικασίες και να τροποποιούν την λειτουργία τους με σκοπό την συνεχή ανάπτυξη, την βελτίωση και τον εκσυγχρονισμό τους.

Η Διοίκηση Επιχειρηματικών Διαδικασιών (Business Process Management) <sup>1</sup> βοηθά τους οργανισμούς να διαχειριστούν τις ανησυχίες που σχετίζονται με τις επιχειρησιακές διαδικασίες <sup>2</sup>. Αναφέρεται τόσο σε διαχειριστικές όσο και σε τεχνικές διαδικασίες οι οποίες έχουν σημαντικό αντίκτυπο στην παραγωγικότητα, την ταχύτητα εξυπηρέτησης των πελατών και βελτίωση υπηρεσιών και προϊόντων. Είναι σημαντικό, επίσης, να επισημανθεί ότι οι επιχειρηματικές διαδικασίες αφορούν τις βασικές λειτουργίες μιας επιχείρησης με σκοπό την υλοποίηση του επιχειρηματικού στόχου και όχι την εργασία την οποία απαιτείται για να υλοποιηθεί αυτός. Τέλος, οι επιχειρηματικές διαδικασίες ταξινομούνται ως ανθρωποκεντρικές και συστηματοκεντρικές διαχωρισμός που γίνεται βάσει το ποιός εκτελεί την διαδικασία και σε ποιόν αναφέρεται.

Οι βασικές δραστηριότητες ενός έργου διοίκησης διαδικασιών περιλαμβάνουν την μοντελοποίηση της διαδικασίας σαν αρχικό σημείο ανάλυσης και παραμετροποίησης, την θέσπιση της διαδικασίας, χρησιμοποιώντας πληροφοριακά συστήματα για να τον έλεγχο και την υποστήριξη της, την ανάλυση της διαδικασίας χρησιμοποιώντας παραγωγικά δεδομένα και μοντέλα διαδικασιών, και τη διαχείριση της διαδικασίας η οποία περιλαμβάνει ένα σύνολο δραστηριοτήτων όπως η προσαρμογή των διαδικασιών, η διαχείριση των πόρων ή η διαχείριση του συνόλου των σχετικών μοντέλων διαδικασιών.

Η σημαντικότητα της Διοίκησης Επιχειρηματικών Διαδικασιών αποδεικνύεται αρ-

---

<sup>1</sup>Η διαχείριση επιχειρησιακών διαδικασιών (Business Process Management) είναι ένα σύνολο μεθόδων, τεχνικών και εργαλείων με στόχο τον εντοπισμό, την ανάλυση, τον ανασχεδιασμό, την εκτέλεση και την παρακολούθηση των επιχειρησιακών διαδικασιών [18].

<sup>2</sup>Επιχειρησιακή διαδικασία ονομάζουμε ένα σύνολο αλληλένδετων γεγονότων, δραστηριοτήτων και στιγμών αποφάσεων που περιλαμβάνουν κάποιους παράγοντες και αντικείμενα και που συλλογικά οδηγούν σε ένα αποτέλεσμα που δίνει αξία τουλάχιστον σε ένα πελάτη[18]

χικά από το μέγεθος της ολοκληρωμένης ακαδημαϊκής ερευνάς η οποία έχει διεξαχθεί τα τελευταία χρόνια εξαιτίας της πρακτικής συνάφειας και των ραγδαίων εξελίξεων που την δικαιολογούν. Αφετέρου από την ανάγκη και δημιουργία ανεξάρτητων τμημάτων σε επιχειρήσεις και οργανισμούς και ,τέλος, από την πληθώρα εργαλείων λογισμικού, τεχνικών και μεθόδων τα οποία υποστηρίζουν τον σχεδιασμό, την θέσπιση και την διαχείριση των επιχειρησιακών διαδικασιών. Πρόκειται για μία υβριδική επιστήμη η οποία συνδυάζει ορολογία, τεχνικές και μεθόδους από τον κλάδο επιστημών διοίκησης και τον κλάδο των επιστημών της πληροφορικής, παράγοντας γνώση η οποία εφαρμόζεται , κυρίως, στην διοίκηση οργανισμών.

## 2.1 Γιατί είναι απαραίτητο το BPM

Η ανάγκη για σαφείς επιχειρηματικές διαδικασίες που να αναλύουν και να καθορίζουν τις επιχειρηματικές δραστηριότητες και τους οικονομικούς στόχους ενός οργανισμού σε μια σειρά από ξεχωριστές, επαναλαμβανόμενες εργασίες με την ελπίδα να επιτύχουν μεγαλύτερη συνέπεια και αξιοπιστία δεν μπορεί να υποτιμηθεί. Ωστόσο, συχνά, στελέχη, ιδιοκτήτες και υπάλληλοι, αναπτύσσουν επιχειρηματικές διαδικασίες χωρίς σαφή αίσθηση πρόβλεψης ή πραγματική κατανόηση του πώς κάθε διαδικασία εντάσσεται στην ευρύτερη οργανωτική δομή της εταιρείας. Συνεπώς, η οργάνωση ενός τμήματος διοίκησης επιχειρηματικών διαδικασιών και η εγκατάσταση των αντίστοιχων πληροφοριακών συστημάτων μπορεί να επιλύσει τέτοιου είδους παθογένειες.

Στις περιπτώσεις που οι εταιρείες υιοθετούν με ορθό και συστηματοποιημένο τρόπο την ΔΕΔ, βιώνουν άφθονα οφέλη στις επιχειρήσεις τους, παρέχοντας μια ολοκληρωμένη κατανόηση του τρόπου λειτουργίας τους με σκοπό να βελτιώσουν την αποδοτικότητα και να αυξήσουν την παραγωγικότητα. Αρχικά, όπως αναφέρθηκε η ΔΕΔ αυξάνει την παραγωγικότητα γιατί δημιουργεί σταθερά πλαίσια συνεργασίας μεταξύ των τμημάτων με αποτέλεσμα οι διαδικασίες να παρακολουθούνται. Συνεπώς η ανάλυση της διαδικασίας γίνεται συνεχώς και ευκολά εντοπίζονται αποκλίσεις, λάθος συμπεριφορές οι οποίες στην συνέχεια εξαλείφονται. Επίσης, με την καθιέρωση, σχεδόν, σε όλους τους οργανισμούς του ευέλικτου τρόπου προγραμματισμού εργασιών (agile, SCRUM, κα.) το BPM κάνει τις ομάδες να είναι πιο ευέλικτες στις αλλαγές γνωρίζοντας και αφομοιώνοντας καλά τις διαδικασίες τους. Όταν υπάρχει πλήρης γνώση των διαδικασιών, είναι εύκολο να κατανοηθεί από τις ομάδες το γιατί και το πώς μιας νέας αλλαγής, ώστε να μπορούν να υιοθετηθούν γρήγορα. Ένας οργανισμός που είναι ευέλικτος στις αλλαγές είναι πολύ ισχυρός. Μπορεί να επιλύσει άμεσα τυχόν προβλήματα που προκαλούνται από εσωτερικούς ή εξωτερικούς παράγοντες και να πετύχει τους επιχειρηματικούς του στόχους. Οι ομάδες γίνονται επεκτάσιμες, ανταγωνιστικές και συνεργάζονται καλύτερα.

Σημαντικό είναι ότι η διοίκηση επιχειρηματικών διαδικασιών εξασφαλίζει την συμμόρφωση με τους κανόνες της εταιρικής πολιτικής, είτε πρόκειται για εσωτερικές είτε για εξωτερικές οδηγίες. Εξορθολογίζει τις διαδικασίες και εφαρμόζει ένα πλαίσιο συμμόρφωσης διασφαλίζοντας ότι οι διαδικασίες είναι καλά κατανοητές και ακολουθούν αυστηρές οδηγίες. Πλέον, υπάρχουν πολλές πολιτικές ανά κλάδο και ανά τμήμα προς συμμόρφωση. Για παράδειγμα, το τμήμα ανθρώπινου δυναμικού έχει διαφορετικούς εσωτερικούς κανόνες προς συμμόρφωση από το τμήμα του λογιστηρίου, αλλά και τα δύο τμήματα συμμορφώνονται με τους γενικούς κανόνες του οργανισμού. Η ΔΕΔ θα πρέπει να διασφαλίζει ότι και οι δύο ομάδες είναι σε θέση να συμμορφώνονται με τις

αντίστοιχες πολιτικές τους στο σύνολο των πληροφοριακών συστημάτων που χρησιμοποιούν.

Τέλος, σημαντικότερο πλεονέκτημα είναι η βελτίωσή των υπηρεσιών και των προϊόντων που φτάνουν στον πελάτη. Όταν οι επιχειρηματικές διαδικασίες είναι ακριβείς και αποτελεσματικές, το προϊόν βελτιώνεται και οδηγεί σε μια καλύτερη συνολική εμπειρία. Στους πελάτες θα παραδοθεί το τελικό προϊόν στους χρόνους που έχουν συμφωνηθεί και χωρίς προβλήματα, έχοντας έτσι μια θετική εμπειρία η οποία συμβάλλει στη συνέχιση της συνεργασίας τους. Επιπλέον, η διαχείριση επιχειρηματικών διαδικασιών μπορεί να βοηθήσει τις επιχειρήσεις να κατανοήσουν καλύτερα τις ανάγκες και τις προτιμήσεις των πελατών τους, επιτρέποντάς τους να δημιουργήσουν προϊόντα και υπηρεσίες που τις καλύπτουν.

## 2.2 Κύκλος Ζωής BPM

Το εύρος των εργασιών της Διοίκησης Επιχειρησιακών Διαδικασιών είναι αρκετά μεγάλο και αφορά σε εργασίες από την αυτοματοποίηση και ανακάλυψη λειτουργιών μέχρι και την βελτίωση οργάνωση και διαχείριση αυτών. Με την χρήση κατάλληλων εργαλείων μπορεί να σχεδιάζει το μοντέλο μιας διαδικασίας, να την προσομοιώνει και, αναλύοντας την, να καταλήξει στον επανασχεδιασμό της και εν τέλει την βελτίωση της. Όπως θα αναλυθεί παρακάτω η έννοια του μοντέλου είναι θεμελιώδης για την Διοίκηση Επιχειρησιακών Διαδικασιών. Πρόκειται για ένα μοντέλο διαδικασίας το οποίο στοχεύει να συλλάβει τους διαφορετικούς τρόπους με τους οποίους μπορεί να διεκπεραιωθεί μια υπόθεση. Υπάρχουν διαφορετικές τεχνικές για να αναπαραστήσουμε μια διαδικασία που σαν κοινό έχουν ότι οι διαδικασίες περιγράφονται με όρους δραστηριοτήτων (BPMN, Petrinet, UML κλπ.) [6]. Η φάση σχεδίασης του μοντέλου διαδικασιών όπως και άλλα σημαντικά σημεία αναφοράς είναι μέρη του κύκλου ζωής Διοίκησης Επιχειρηματικών Διαδικασιών (BPM life cycle).

Σε γενικές γραμμές, ο κύκλος ζωής ενός έργου Διοίκησης Επιχειρηματικών Διαδικασιών αποτελείται από τρεις φάσεις, οι οποίες αντίστοιχα βασίζονται σε συγκεκριμένες τεχνικές υλοποίησης και ανάλυσης. Στην πρώτη φάση του σχεδιασμού/επανασχεδιασμού γίνεται ο αρχικός σχεδιασμός του μοντέλου βάσει των προδιαγραφών της διαδικασίας, στην δεύτερη φάση υλοποίησης και παραμετροποίησης η διαδικασία μετατρέπεται σε ένα λειτουργικό σύστημα, φάση η οποία χρονικά είναι συντομότερη αν η διαδικασία βρίσκεται ήδη σε παραγωγική λειτουργία και τέλος, η τρίτη φάση εκτέλεσης και εφαρμογής στην οποία οι διαδικασίες είναι πλέον παραγωγικές. Στην τρίτη φάση η διαδικασία δεν επανασχεδιάζεται ούτε δημιουργείται κάποιο νέο λογισμικό, αλλά γίνεται εξόρυξη δεδομένων με στόχο την μετέπειτα αξιολόγηση και τον ενδεχόμενο επανασχεδιασμό της, αν αυτό κρίνεται απαραίτητο βάσει της ανάλυσης. Όπως θα αναφερθεί παρακάτω από τα δεδομένα που θα εξορυχθούν κατά την φάση λειτουργίας της διαδικασίας με την κατάλληλη χρήση αλγορίθμων ανακάλυψης διαδικασιών, γίνεται ανακάλυψη αποκλίσεων (deviation), συμφορήσεων (bottlenecks) και περιττών ενεργειών (wastes). Παράλληλα, στο υπάρχον μοντέλο μπορεί να γίνει μια προσομοίωση με παραγωγικά δεδομένα, μια εις βάθος “what-if” ανάλυση ή να επαληθευτεί η ορθότητα ενός μοντέλου της διαδικασίας [6].

Για την καλύτερη κατανόηση των εννοιών, των τεχνολογιών που χρησιμοποιεί η κάθε φάση και των λογικών εξαρτήσεων των επιμέρους φάσεων μεταξύ τους, θα πρέπει να γίνει αναφορά στα στάδια που τις αποτελούν. Αναλυτικότερα την φάση σχεδιασμού, μπορούμε να την διασπάσουμε σε επιμέρους εργασίες:

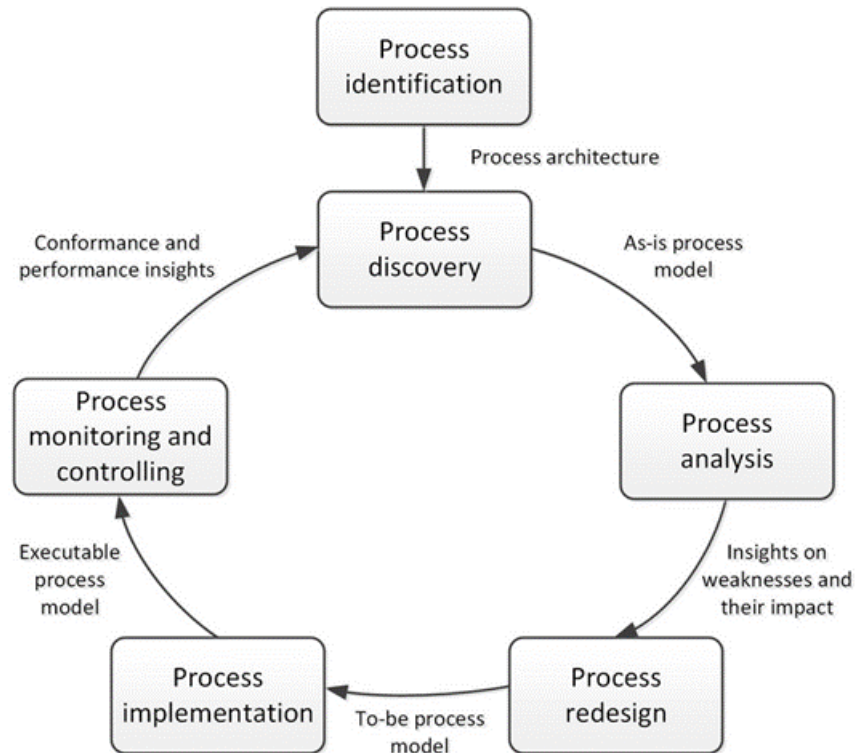


Figure 2.1: Αναλυτικός κύκλος ζωής BPM. Πηγή[18]

- *Προσδιορισμός εργασίας.* Σε αυτό το στάδιο ο οργανισμός θέτει έναν επιχειρησιακό στόχο τον οποίο θέλει να προσθέσει στην λειτουργία του. Αναλογικά,= οι στόχοι ενός οργανισμού διακρίνονται σε οικονομικούς, πελατοκεντρικούς, εσωτερικών λειτουργιών και διαδικασιών και στόχους ανάπτυξης ανθρώπινου δυναμικού. Το στάδιο αυτό συμπεριλαμβάνεται στην δημιουργία μια νέας διαδικασίας
- *Εντοπισμός διαδικασίας.* Στο στάδιο του εντοπισμού της διαδικασίας, βάσει ερευνάς, καθορίζεται το εννοιολογικό πλαίσιο λειτουργίας της διαδικασίας και πραγματοποιείται η μοντελοποίηση της διαδικασίας. Ένα μοντέλο διαδικασιών μπορεί περιγράφει την σειρά των δραστηριοτήτων, τις ιδιότητες τους, όπως ο χρόνος και οι πόροι, και τις συσχετίσεις μεταξύ τους. Εκτός από τις διαφορετικές τεχνικές μοντελοποίησης υπάρχουν και οι διαφορετικές οπτικές. Έτσι ένα μοντέλο μπορεί να περιγράφει βάσει της ροής δραστηριοτήτων(control-flow perspective), των χρονικών τους ιδιοτήτων(time perspective), την δημιουργία και χρήση δεδομένων(data perspective), την αξιοποίηση των πόρων και το τρόπο με τον οποίο αλληλοεπιδρούν στην διαδικασία(resource perspective) και των σχέσεων μεταξύ των δραστηριοτήτων(function perspective). Τα μοντέλα τα οποία προκύπτουν χρησιμοποιούνται συνήθως για την κατασκευή ή την διαμόρφωση πληροφοριακών συστημάτων καθώς και για την περαιτέρω ανάλυση, κατανόηση και βελτίωση των διαδικασιών που περιγράφουν. Επίσης, η μοντελοποίηση επιχειρηματικών διαδικασιών είναι μία εξελικτική εργασία υπό την έννοια ότι το μοντέλο της διαδικασίας αναλύεται και βελτιώνεται έτσι ώστε να οπτικοποιεί το επιθυμητό αποτέλεσμα χωρίς να περιέχει ανεπιθύμητες συμπεριφορές.

- *Ανάλυση διαδικασίας.* Στο στάδιο της ανάλυσης γίνεται η εφαρμογή περιπτώσεων χρήσης στο μοντέλο διαδικασίας που δημιουργήθηκε, μετρίεται η διαδικασία ποιοτικά και ποσοτικά με σκοπό την καταγραφή και τον προσδιορισμό των προβλημάτων που τυχόν προκύψουν. Γίνεται προσομοίωση με την χρήση διαφορετικών συνδυασμών μεταβλητών και δεδομένων. Το αποτέλεσμα αυτής της φάσης είναι μια σειρά από προτάσεις για τροποποιήσεις του υπάρχοντος μοντέλου.
- *Βελτίωση διαδικασίας.* Στο στάδιο αυτό προκύπτει το τελικό μοντέλο διαδικασίας με βάση τις προτάσεις του βήματος της ανάλυσης που θα αποτελέσει βάση της υλοποίησης της διαδικασίας. Σε αυτό το μοντέλο όλα τα προβλήματα έχουν επιλυθεί αλλά ταυτόχρονα θα πρέπει να υπολογιστεί επίσης, το κατά πόσο είναι εφικτό να γίνουν οι αλλαγές. Δηλαδή, αν πρόκειται το μοντέλο να αλλάξει σε μεγάλο βαθμό πρέπει να υπολογιστεί το αντίστοιχο κόστος και ο χρόνος που απαιτούνται για την εφαρμογή τους.

Επόμενο στάδιο είναι αυτό της υλοποίησης και της παραμετροποίησης της διαδικασίας το οποίο αναλυτικότερα περιλαμβάνει τα παρακάτω στάδια.

- *Υλοποίηση της διαδικασίας.* Πρόκειται για το βήμα υλοποίησης της διαδικασίας σε παραγωγικό περιβάλλον. Σταδιακά μπαίνει σε εφαρμογή, αρχικά εξυπηρετώντας ένα μικρό μέρος του οργανισμού και, αφού εκτελεστούν αρκετοί κύκλοι δοκιμών και η διαδικασία λειτουργεί χωρίς προβλήματα βάσει του μοντέλου διαδικασίας που έχει αναπτυχθεί, τότε μπαίνει σε πλήρη λειτουργία από τον οργανισμό. Σε αυτό το στάδιο και αλλά τμήματα του οργανισμού ενεργοποιούνται και βοηθούν στην υλοποίηση, το τμήμα πληροφορικής εγκαθιστά το απαραίτητο λογισμικό και οι άνθρωποι του οργανισμού πρέπει να εκπαιδευτούν στην νέα διαδικασία ώστε να χρησιμοποιούν ορθά το σύστημα. Όταν, ειδικά, πρόκειται για μεγάλη αλλαγή στην διαδικασία είναι αναγκαίο να κατανοήσει ολόκληρος ο οργανισμός τα πλεονεκτήματα της χρησιμότητάς της, τομέας που αναφέρεται σε άλλο επιστημονικό πεδίο (change management) που δεν αφορά το θέμα της παρούσας διπλωματικής. Σε διαδικασίες οι οποίες είναι σε λειτουργία και ο οργανισμός είναι ώριμος με την εκτέλεση τους και πρόκειται για βελτίωση τους το στάδιο αυτό είναι αρκετά σύντομο.

Τελευταίο στάδιο είναι το στάδιο της εφαρμογής και παρακολούθησης της διαδικασίας. Περιλαμβάνει τα παρακάτω στάδια:

- *Έλεγχος της διαδικασίας.* Η διαδικασία πλέον βρίσκεται σε παραγωγικό στάδιο εξυπηρετώντας τις ανάγκες και τους επιχειρηματικούς σκοπούς του οργανισμού. Κύριος στόχος είναι η παρακολούθηση και η συνεχής βελτίωση των χρόνων διεκπεραίωσης, της αποδοτικότητας κόστους, των ποσοστών σφαλμάτων και άλλων βασικών δεικτών απόδοσης, για παράδειγμα με την αυτοματοποίηση ή τη σύνδεση διαδικασιών. Κατά την παρακολούθηση παράγονται δεδομένα τα οποία τα οποία, αν αναλυθούν περαιτέρω, συμβάλουν στον εντοπισμό ενδεχόμενων προβλημάτων τα οποία υπάρχουν (αποκλίσεις, λάθος συμπεριφορές κ.α.).
- *Ανάλυση διαδικασίας.* Με βάση την ανάλυση των δεδομένων που ανακτήθηκαν, στη συνέχεια προτείνονται μέτρα βελτίωσης σχετικά με τις διαδικασίες και τους ορισμούς στόχων που εφαρμόζονται. Μέρος αυτής της φάσης είναι και η εξόρυξη διαδικασιών (Process Mining). Με την χρήση αλγόριθμων ανακάλυψης διαδικασιών και την εφαρμογή των τεχνικών και τεχνολογιών που τους διέπουν, τα δεδομένα αποκτούν αξία και οργανισμός είναι σε θέση να βελτιώσει την λειτουργία του.



## 2.3 Συστήματα BPM

Η διοίκηση επιχειρηματικών διαδικασιών στην εποχή της τέταρτης βιομηχανικής επανάστασης, δηλαδή την εποχή της αυτοματοποίησης και της ανταλλαγής δεδομένων στις τεχνολογίες παραγωγής, σχετίζεται, ως επί το πλείστον, με συστήματα λογισμικού με τα οποία διαχειρίζεται, ελέγχει και υποστηρίζει ένα έργο διοίκησης επιχειρηματικών διαδικασιών. Αποτελεί το «τρίτο κύμα» στον κλάδο της μηχανικής διαδικασιών<sup>3</sup> και, συγκεκριμένα, αποτελεί μέλος της εξέλιξης σε μηχανική συστημάτων διαδικασιώνίως βάλω ορισμό. Το πρώτο κύμα αφορούσε «έγγραφα διαδικασίας» που αναδιοργάνωναν και καθόριζαν την ανθρώπινη δραστηριότητά, το δεύτερο κύμα επικεντρώθηκε στον ανασχεδιασμό των επιχειρηματικών διαδικασιών με την χρήση ERP (Enterprise resource planning) και το τρίτο εστιάζει σε επιχειρηματικά μοντέλα διαδικασιών και την ικανότητα τους να παράγουν δεδομένα ώστε να έχουν την δυνατότητα να τροποποιηθούν, βελτιωθούν και εναρμονιστούν με τις εκάστοτε ανάγκες του οργανισμού[26].

Τις πρώτες δεκαετίες που οι επιχειρήσεις άρχισαν να χρησιμοποιούν πληροφοριακά συστήματα για τις λειτουργίες τους, οι εφαρμογές είχαν την μορφή ενιαίου προγράμματος. Η ανάγκη και αποθήκευση των δεδομένων ήταν, επίσης, ενσωματωμένη σε αυτό. Με την πάροδο του χρόνου η ανάγκη για την κοινή διαχείριση των δεδομένων από διαφορετικά συστήματα προσάρτησε στις εφαρμογές σχεσιακές βάσεις δεδομένων (DBMS). Στην συνέχεια, η ανάπτυξη των λειτουργικών συστημάτων (Windows, Linux, MACOS κα.) και η προσθήκη σε αυτά λειτουργικότητας που αφορά την αλληλεπίδρασή τους με τον χρήστη ενίσχυσε τις εφαρμογές και τα πληροφοριακά συστήματα με γραφικό περιβάλλον. Τέλος, η τάση για τυποποίηση και επαναλαμβανόμενη λειτουργία κάποιων διαδικασιών της επιχείρησης δημιούργησε την ανάγκη προσθήκης ενός συστήματος το οποίο να φροντίζει για την σωστή διεκπεραίωση των εργασιών ώστε η εφαρμογή να επικεντρώνεται σε συγκεκριμένες λειτουργίες και να είναι πιο αποδοτική. Η αρχή έγινε με τα συστήματα διαχείρισης εργασιών (WFM), τα οποία επικεντρωνόντουσαν στην αυτοματοποίηση της εργασίας χωρίς να υποστηρίζουν ανάλυση των δραστηριοτήτων που εκτελούσαν αλλά ούτε και την διαχείριση τους. Για αυτό και στην συνέχεια αντικαταστάθηκαν από συστήματα διαχείρισης επιχειρηματικών διαδικασιών, συστήματα που υποστηρίζουν κάθε αυτού τις διαδικασίες. Παρέχουν λειτουργίες όπως η προσομοίωση, η ευφυΐα των επιχειρηματικών διαδικασιών (process intelligence), διαχείριση υποθέσεων (case management), κ.α. [18].

Η εξέλιξη επιτρέπει πλέον, στα πληροφοριακά συστήματα να έχουν γνώση για το πώς πρέπει να γίνονται οι διαδικασίες παρέχοντας μεγάλη ευελιξία στην καταγραφή, έλεγχο και υποστήριξη των διαδικασιών. Περαιτέρω διαχωρισμός μπορεί να γίνει σε συστήματα τα οποία γνωρίζουν την διαδικασία (Process Aware Information

---

<sup>3</sup>Η μηχανική διεργασιών (Process Engineering) είναι η κατανόηση και η εφαρμογή των θεμελιωδών αρχών και νόμων της φύσης που επιτρέπουν στους ανθρώπους να μετατρέψουν την πρώτη ύλη και την ενέργεια σε προϊόντα χρήσιμα για την κοινωνία, σε βιομηχανικό επίπεδο. Εκμεταλλευόμενοι τις κινητήριες δυνάμεις της φύσης όπως η πίεση, η θερμοκρασία και οι διαβαθμίσεις συγκέντρωσης, καθώς και ο νόμος της διατήρησης της μάζας, οι μηχανικοί διεργασιών μπορούν να αναπτύξουν μεθόδους για τη σύνθεση και τον καθαρισμό μεγάλων ποσοτήτων επιθυμητών χημικών προϊόντων. Η μηχανική διεργασιών επικεντρώνεται στο σχεδιασμό, τη λειτουργία, τον έλεγχο, τη βελτιστοποίηση και την εντατικοποίηση χημικών, φυσικών και βιολογικών διεργασιών. Η μηχανική διεργασιών περιλαμβάνει ένα ευρύ φάσμα βιομηχανιών, όπως η γεωργία, η αυτοκινητοβιομηχανία, η βιοτεχνική, η χημική, τα τρόφιμα, η ανάπτυξη υλικών, η εξόρυξη, η πυρηνική, η πετροχημική, η φαρμακευτική και η ανάπτυξη λογισμικού. Η εφαρμογή συστηματικών μεθόδων που βασίζονται σε υπολογιστή στη μηχανική διεργασιών είναι η «μηχανική συστημάτων διεργασιών». Wikipedia

Systems, PAISs) και σε συστήματα τα οποία υποστηρίζουν την διαδικασία αλλά δεν έχουν γνώση το πως θα γίνει(non Process Aware Information Systems). Μερικά παραδείγματα PAISs συστημάτων είναι τα συστήματα διοίκησης επιχειρήσεων(BPM systems), συστήματα διαχείρισης εργασιών(WFM systems), συστήματα διαχείρισης πελατών(CRM systems), συστήματα διαχείρισης οργανισμών(ERP systems). Παραδείγματα των non-PAISs συστημάτων είναι οι βάσεις δεδομένων, ένας mail server κα. Κοινό χαρακτηριστικό των δύο αυτών συστημάτων είναι ότι με την κατάλληλη επεξεργασία και τα εργαλεία μπορούν να αναγνωρίσουν διαδικασίες και να εφαρμόσουν BPM τεχνικές, όπως η εξόρυξη διαδικασιών(Process Mining).



# Εξόρυξη Διαδικασιών

Η εξόρυξη διαδικασιών είναι ένα σύνολο τεχνικών και εργαλείων το οποίο έχει σαν είσοδο ένα επεξεργασμένα δεδομένα, τα οποία έχουν προκύψει από τα πληροφοριακά συστήματα που υπάρχουν και εξυπηρετούν τις ανάγκες του οργανισμού, με σκοπό να ανακαλύψουν τις παραγωγικές διαδικασίες και να τις βελτιώσουν, υποστηρίζοντας παράλληλα την Διοίκηση Επιχειρηματικών Διαδικασιών. Η επιστήμη της εξόρυξης διαδικασιών εκμεταλλεύεται στο έπακρο ότι όλοι οι οργανισμοί διαθέτουν πληροφοριακά συστήματα, όπως CRM, ERP και HCM, για την υποστήριξη της συνολικής λειτουργίας τους και ταυτόχρονα την καταγραφή δεδομένων σχετικά με τις διαδικασίες τις οποίες υποστηρίζουν.

Με αυτόν τον τρόπο οι αναλυτές αποκτούν πραγματική αντίληψη για το τί γίνεται πραγματικά στο εσωτερικό του οργανισμού, οι ενδεχόμενες αστοχίες μπορούν να αναλυθούν και η ποιότητα των μοντέλων διαδικασιών μπορεί να βελτιωθεί βάσει συγκεκριμένων επιχειρησιακών κριτηρίων, όπως το κόστος, ο χρόνος και η ποιότητα της διαδικασίας. Στο τέλος δε, υπάρχει η δυνατότητα να «κλείσει» ο κύκλος ζωής ενός έργου διοίκησης επιχειρηματικών διαδικασιών[21] και να δημιουργηθεί εκ νέου μόνο όταν υπάρξει ανάγκη για ριζικές αλλαγές στις διαδικασίες. Όχι βελτιώσεις στις ήδη υπάρχουσες παραγωγικές.

Η εξόρυξη διαδικασιών πλεονεκτεί από άλλα αντίστοιχα έργα λήψης αποφάσεων και έργα διαχείρισης επιχειρηματικών διαδικασιών, αρχικά ως προς την πληθώρα των διαθέσιμων αλγόριθμων ανακάλυψης διαδικασιών. Έτσι παροτρύνει τους οργανισμούς να πειραματιστούν και να επιλέξουν τον κατάλληλο αλγόριθμο αντίστοιχό ανάλογα με την ποιότητά των δεδομένων που έχουν εξορύξει και τους στόχους ανάλυσης που έχουν θέσει για τη διαδικασία. Εφόσον, τα αποτελέσματα ενδέχεται να ποικίλουν ανάλογα με την ποιότητα των δεδομένων και δεν περιορίζονται σε βασικές αρχές, μπορεί να γίνει ανακάλυψη των διαδικασιών με την ελάχιστη πληροφορία που έχει οριστεί σαν υποχρεωτική. Παράλληλα όμως μπορούν τα αρχεία καταγραφής να εμπλουτιστούν με περισσότερες ιδιότητες για λεπτομερή ανάλυση και εξαγωγή διαφορετικού μοντέλου διαδικασίας. Τέλος, η δυνατότητα της οπτικοποίησης των αποτελέσματα της εξόρυξης διαδικασιών, της παρουσίας με ποικίλους τρόπους (PetriNets, BPMN, οργανωτικούς χάρτες, δένδρα διαδικασιών), της ανάλυσης με την χρήση τεχνικών από διαφορετικούς τομείς (data mining) και της ευελιξία που έχουν σε επίπεδο ευκρινείας την κατατάσσουν σε μια προσοδοφόρα τεχνική που μπορεί να εφαρμοστεί στα πλαίσια της λήψης αποφάσεων του οργανισμού[2].

Ο κλάδος της εξόρυξης διαδικασιών αποτελεί μια γέφυρα μεταξύ του τομέα της εξόρυξης δεδομένων (Data Mining) και την αντίστοιχη επιστήμη μελέτης των διαδικασιών[21]. Η γέφυρά αυτή αποδεικνύεται αρχικά από το γεγονός ότι η εξόρυξη διαδικασιών έχει αναδειχθεί σε μία πρωτοπόρο τεχνολογία αναφορικά με την ανάλυση Μεγάλων Δεδομένων (Big Data) [2] και, έπειτα, από την αποδεδειγμένη εφαρμογή των τεχνικών

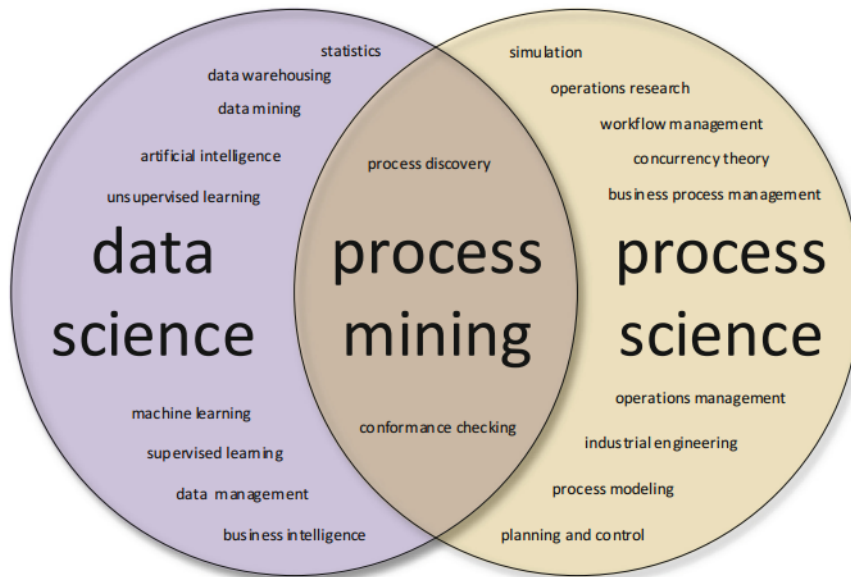


Figure 3.1: To Process Mining, γέφυρά μεταξύ του Data Mining και του Process Management. Πηγή[21]

που εφαρμόζονται πάνω σε αυτά τα δεδομένα προσθέτει αξία και παράγει κέρδος για τον οργανισμό. Η αξία που παράγεται σύμφωνα με βιβλιογραφικές έρευνες [2] είναι, αρχικά η βελτίωση της ποιότητας των δεδομένων, δηλαδή των αρχείων καταγραφής τα οποία πολλές φορές στις αρχικές προσπάθειες εξόρυξης της διαδικασίας κρίνονται ανεπαρκή, θα οδηγήσει στην ανάπτυξη νέων τεχνολογικών δυνατοτήτων του οργανισμού ενισχύοντας την καινοτομία. Επιπλέον, οι τεχνικές εξόρυξης διαδικασιών ενισχύουν και διευκολύνουν τις οργανωτικές πρακτικές και συγκεκριμένα την ανάγκη που υπάρχει για τον σχεδιασμό ευρέων ολιστικών διαδικασιών εντάσσοντας τις στο γενικό εννοιολογικό πλαίσιο του οργανισμού. Αυτό συμβαίνει γιατί, πρώτον, η εξόρυξη διαδικασιών αποδεδειγμένα[2] ενισχύει τη διαφάνεια σε όλο το φάσμα των διαδικασιών και διευκολύνει την παρακολούθηση του ανασχεδιασμού τους και, δεύτερον, η άμεση επιρροή των εργαζομένων με τα αποτελέσματα της εξόρυξης διαδικασιών ενισχύει την «προσανατολισμένη στη διαδικασία σκέψη» τους γνωρίζοντας το πώς οι επιχειρηματικές διαδικασίες εκτελούνται και αλληλοσυνδέονται στην πραγματικότητα[2]. Τέλος, υπάρχουν διαφανείς διαδικασίες. Με την χρήση πραγματικών δεδομένων, τα στελέχη των οργανισμών μπορούν να έχουν καλύτερο έλεγχο, οι αποφάσεις που λαμβάνονται να είναι άμεσα υλοποιήσιμες, τα προβλήματα και οι καθυστερήσεις να εντοπίζονται άμεσα.

Επιπλέον, τα αποτελέσματα των ερευνών [15] που πραγματοποιήθηκαν σε οργανισμούς και εταιρείες που εφαρμόζουν τεχνολογίες εξόρυξης διαδικασιών έδειξαν ότι η αξία προκύπτει από τα παρακάτω χαρακτηριστικά: πρώτον, από την αντίληψη που δημιουργείται μέσω της οπτικοποίησης του συνόλου της διαδικασίας και την μέτρηση απόδοσης της, δεύτερον, από την κατανόηση των σχετικών με την διαδικασία πληροφοριών, τρίτον, από την λήψη αποφάσεων βάσει των δεδομένων και, τέταρτον, από τη υλοποίηση και εφαρμογή παρεμβάσεων σε αυτήν. Τέλος, πρέπει να γίνει αντιληπτό ότι η αξία που προστίθεται δεν αφορά το κέρδος σε σχέση με τον προϋπολογισμό αλλά και με την αξία που δίνει στις εσωτερικές διαδικασίες του οργανισμού και την ενίσχυση της συνεργασίας, της ευεξίας των εργαζομένων, της διαφάνειας του φόρτου εργασίας κ.α.

### 3.1 Event Logs

Θεμελιώδης έννοια της επιστήμης εξόρυξης διαδικασιών είναι τα αρχεία καταγραφής συμβάντων (event logs)<sup>1</sup> των οποίων η εξόρυξη γίνεται μέσα από συστήματα καταγραφής δεδομένων όπως είναι οι βάσεις δεδομένων, τα αρχεία καταγραφής συναλλαγών, αρχεία καταγραφής μηνυμάτων, επιχειρησιακά συστήματα ERP, CRM, HCM κ.α. Ένα αρχείο καταγραφής συμβάντων είναι μια χρονολογική καταγραφή γεγονότων που συμβαίνουν κατά την διάρκεια εκτέλεσης μιας διαδικασίας. Στο πλαίσιο της εξόρυξης διεργασιών, ένα αρχείο καταγραφής συμβάντων είναι η αντίστοιχη ψηφιακή εγγραφή των δραστηριοτήτων για την ολοκλήρωση μιας επιχειρηματικής διαδικασίας. Τα αρχεία καταγραφής συμβάντων καταγράφουν πληροφορίες όπως η ώρα που έλαβε χώρα μια συγκεκριμένη δραστηριότητα, το άτομο ή το σύστημα που πραγματοποίησε τη δραστηριότητα και το αποτέλεσμα της δραστηριότητας. Η επιστήμη εξόρυξης διαδικασιών χρησιμοποιώντας τα αρχεία καταγραφής συμβάντων μέσα από την οπτική των διαδικασιών μπορεί να απαντήσει σε επιχειρησιακές ερωτήσεις σχετικά με την διαδικασία που αναλύει. Ερωτήσεις όπως, πώς και το γιατί έλαβε χώρα μια συγκεκριμένη συμπεριφορά στο παρελθόν, τί μπορεί να συμβεί στο μέλλον, γιατί ο οργανισμός και οι εργαζόμενοι σε αυτόν έχουν αποκλίνουσες συμπεριφορές, τί ενέργειες πρέπει να γίνουν για να βελτιωθεί η υπάρχουσα διαδικασία.

Τα αρχεία καταγραφής συμβάντων διαφοροποιούνται αρκετά με τα δεδομένα που χρησιμοποιούνται για τις παραδοσιακές τεχνικές ανάλυσης δεδομένων. Αποτελούνται από ένα σύνολο περιπτώσεων χρήσης οι οποίες περιέχουν ένα αριθμό από γεγονότα τα οποία αντιστοιχούν σε μια δραστηριότητα, συσχετίζονται μεταξύ τους βάσει χρονολογικής σειράς, μιας γενικής εικόνας του συνόλου της διαδικασίας. Τα γεγονότα περιέχουν ένα σύνολο ιδιοτήτων τις οποίες ένα έργο εξόρυξης διαδικασιών μπορεί να χρησιμοποιήσει για να εμπλουτίσει την διαδικασία και να δημιουργήσει λεπτομερέστερα μοντέλα διαδικασιών. Έχει θεσπιστεί ένας ελάχιστος αριθμός χαρακτηριστικών, με σκοπό τα αρχεία καταγραφής δεδομένων να απεικονίσουν την ροή μια επιχειρηματικής διαδικασίας από την οποία μπορεί να παραχθεί γνώση μέσω των τεχνικών εξόρυξης δεδομένων, χαρακτηριστικά τα οποία περιγράφονται παρακάτω.

- Το "*Case ID*", πρόκειται για μία μοναδική τιμή σε κάποιο χαρακτηριστικό του στιγμιότυπου της διαδικασίας (Process Instance<sup>2</sup>). Έτσι κάθε γεγονός (event) συνδέεται άμεσα με συγκεκριμένη περίπτωση, με αποτέλεσμά ένα σύνολο από γεγονότα με το ίδιο "Case ID" να αποτελούν το στιγμιότυπο διαδικασίας με σκοπό να αναλυθεί κάθε περίπτωση ανεξάρτητα και να κατανοηθεί καλύτερα η συνολική απόδοση της διαδικασίας. Είναι ιδιαίτερα χρήσιμο στην ανάλυση πολύπλοκων συστημάτων που περιλαμβάνουν περιπτώσεις που εκτελούνται ταυτόχρονα.

<sup>1</sup>Τα αρχεία καταγραφής συμβάντων αποθηκεύουν πληροφορίες σχετικά με το ποιες δραστηριότητες (δηλαδή, το όνομα τη δραστηριότητας) που εκτελούνται για ένα συγκεκριμένο επιχειρηματικό αντικείμενο (δηλαδή, το αναγνωριστικό του επιχειρηματικού αντικειμένου) και το πότε αυτές εκτελούνται (δηλαδή χρονική στιγμή). Μπορούν επίσης να περιέχουν περαιτέρω πληροφορίες, όπως ποιος πραγματοποίησε τη δραστηριότητα (δηλαδή, οι πόροι) και κάτω από ποιες συνθήκες πραγματοποιήθηκε (δηλαδή σύστημα εργασίας).[21]

<sup>2</sup>Το στιγμιότυπο διαδικασίας αναφέρεται σε μια συγκεκριμένη εκτέλεση μιας επιχειρηματικής διαδικασίας από την αρχή μέχρι το τέλος, συμπεριλαμβανομένων όλων των δραστηριοτήτων και αποφάσεων που λαμβάνονται στο πλαίσιο της διαδικασίας. Είναι μια αναπαράσταση μιας μεμονωμένης εμφάνισης μιας διαδικασίας και χαρακτηρίζεται από ένα μοναδικό αναγνωριστικό περίπτωσης που επιτρέπει τη διάκρισή της από άλλες περιπτώσεις. [21]

	Case ID	Timestamp	Medium	Activity	Service Line	Urgency
1	CaseID	Timestamp	Medium	Activity	Service Line	Urgency
2	case9700	20.8.09 11:46	Phone	Registered	1st line	0
3	case9700	20.8.09 11:50	Phone	Completed	1st line	0
4	case9701	23.9.09 12:23	Phone	Registered	1st line	0
5	case9701	23.9.09 12:27	Phone	Completed	1st line	0
6	case9705	20.10.09 14:21	Phone	Registered	Specialist	2
7	case9705	20.10.09 16:48	Phone	At specialist	Specialist	2
8	case9705	19.11.09 10:31	Phone	In progress	Specialist	2
9	case9705	19.11.09 10:32	Phone	Completed	Specialist	2
10	case3939	15.10.09 11:48	Mail	Registered	Specialist	2
11	case3939	15.10.09 11:48	Mail	Offered	Specialist	2
12	case3939	20.10.09 17:18	Mail	In progress	Specialist	2
13	case3939	20.10.09 17:19	Mail	At specialist	Specialist	2
14	case3939	21.10.09 14:49	Mail	In progress	Specialist	2
15	case3939	21.10.09 14:49	Mail	In progress	Specialist	2
16	case3939	28.10.09 10:17	Mail	In progress	Specialist	2
17	case3939	28.10.09 10:18	Mail	Completed	Specialist	2
18	case9704	20.10.09 14:19	Mail	Registered	1st line	0
19	case9704	20.10.09 14:24	Mail	Completed	1st line	0
20	case9703	20.10.09 14:40	Phone	Registered	1st line	0
21	case9703	20.10.09 14:58	Phone	Completed	1st line	0
22	case9702	24.8.09 12:24	Mail	Registered	2nd line	2
23	case9702	24.8.09 12:30	Mail	Offered	2nd line	2

Figure 3.2: Παράδειγμα Αρχείου καταγραφής συμβάντων. Πηγή[1]

- Το "*Activity Name*", με το οποίο κάθε γεγονός αντιστοιχίζεται σε μια συγκεκριμένη δραστηριότητα η οποία αποτελεί μέρος του συνόλου της επιχειρηματικής διαδικασίας. Είναι απαραίτητο για τον προσδιορισμό των συγκεκριμένων βημάτων που έγιναν κατά την διάρκεια εκτέλεσης της διαδικασίας και για την κατανόηση του τρόπου που δίνουν αξία στο τελικό αποτέλεσμα.
- Το "*TimeStamp*", τελευταία προϋπόθεση που υποδεικνύει την χρονική στιγμή την οποία έγινε μια δραστηριότητα της επιχειρηματικής διαδικασίας. Συνδέεται με την ανάγκη να τοποθετηθούν τα γεγονότα που αφορούν μία περίπτωση σε χρονολογική διάταξη και κατά την ανάλυση της απόδοσης της διαδικασίας, εντοπίζοντας τα σημεία συμφόρησης και ενδεχόμενες καθυστερήσεις.

Εάν το ενδιαφέρον ενός έργου εξόρυξης διαδικασιών είναι μόνο ο έλεγχος της ροής και οι εξαρτήσεις μεταξύ των δραστηριοτήτων, τα χαρακτηριστικά που αναφέρθηκαν είναι αρκετά για την εξαγωγή συμπερασμάτων. Σε μία ρεαλιστική παραγωγική περίπτωση και εξαιτίας του γεγονότος ότι ένας οργανισμός χρειάζεται να βγάλει χρήσιμα συμπεράσματα, να βελτιώσει την διαδικασία μέσω τεχνικών εξόρυξης διαδικασιών όπως είναι η βελτιστοποίηση (enhancement) και να μπορέσει να ανακαλύψει μοντέλα διαδικασιών που να δείχνουν την διαδικασία από διαφορετικές οπτικές, πρέπει τα αρχεία καταγραφής να περιέχουν πρόσθετα χαρακτηριστικά. Τα επιπλέον δεδομένα των αρχείων καταγραφής γεγονότων μπορούν να χρησιμεύσουν σε φιλτράρισμα περιπτώσεων και γεγονότων, προκειμένου να επιτευχθεί μια πιο εστιασμένη ανάλυση της διαδικασίας ή να πραγματοποιηθεί συγκριτική ανάλυση μεταξύ υποπεριπτώσεων. Επίσης, ενδέχεται να περιέχουν πολύτιμες πληροφορίες σχετικά με το εννοιολογικό πλαίσιο. Επομένως, μπορούν να αξιοποιηθούν για την απόκτηση καλύτερης αντίληψης σχετικά με τη διαδικασία[20]. Μερικά χαρακτηριστικά περιγράφονται παρακάτω.

- Τα "*Resources*", οι πόροι αναφέρονται σε άτομα ή το πληροφοριακό σύστημα που πραγματοποίησε τη δραστηριότητα. Το χαρακτηριστικό αυτό είναι απαραίτητο για την κατανόηση του τρόπου με τον οποίο κατανέμονται οι ευθύνες στη διαδικασία

και για τον εντοπισμό πιθανών ζητημάτων που σχετίζονται με το φόρτο εργασίας ή το επίπεδο δεξιοτήτων. Με το συνδυασμό επιπλέον χαρακτηριστικών γίνεται η εξόρυξη διαδικασιών από την οπτική της οργανωσιακής δομής του οργανισμού (organizational perspective<sup>3</sup>).

- Το "*Outcome*", αναφέρεται στο αποτέλεσμα μιας δραστηριότητας εντός της διαδικασίας, όπως εάν ήταν επιτυχής ή όχι. Αυτό το χαρακτηριστικό χρησιμοποιείται για τον εντοπισμό αναποτελεσματικών διαδικασιών και σφαλμάτων που προκαλούν αποκλίσεις ή συμφορήσεις εξαιτίας της μη ορθής ολοκλήρωσης.
- Το "*Event Status*", αναφέρεται στην κατάσταση στην οποία βρίσκεται η διαδικασία. Σε ένα event log ενδέχεται μία δραστηριότητα να έχει ίδιο όνομα αλλά να διαφέρει ως προς την κατάσταση. Δείχνει επίσης στο στάδιο που βρίσκεται μια δραστηριότητα.
- Το "*Event Cost*", που αναφέρεται στο κόστος που σχετίζεται με μια δραστηριότητα εντός της διαδικασίας. Χρησιμεύει για τον εντοπισμό περιοχών όπου το κόστος μπορεί να μειωθεί ή να βελτιστοποιηθεί και να βοηθήσει στη συνολική αποτελεσματικότητα της διαδικασίας.

Η εξόρυξη των αρχείων καταγραφής συμβάντων εξαιτίας, της ευρέως διαδεδομένης και δραστηρικής αύξησης της διαθεσιμότητας τέτοιου είδους δεδομένων, έχει γίνει αρκετά πολύπλοκη και δύσκολη. Άλλωστε, η ποιότητα των δεδομένων αποτελεί ένα από τα σημαντικότερα μειονεκτήματα του τομέα εξόρυξης διαδικασιών και ένα πεδίο το οποίο προσφέρεται για εργασίες βελτίωσης. Η έλλειψη ποιότητας δεδομένων μπορεί να οδηγήσει σε ανακριβή ή ελλιπή μοντέλα διαδικασίας, τα οποία ενδέχεται να επηρεάσουν την ακρίβεια των πληροφοριών που αποκτήθηκαν. Τα δεδομένα προέρχονται συνήθως από τα πληροφοριακά συστήματα εταιρειών στον τομέα υπηρεσιών, συμπεριλαμβανομένων τραπεζών και ασφαλιστικών εταιρειών, εταιρείες πωλήσεων, στην υγειονομική περίθαλψη και την εκπαίδευση[22][24]. Μια κατηγοριοποίηση των πληροφοριακών συστημάτων στα οποία έχει γίνει αναφορά και σε προηγούμενα κεφάλαια και ανήκουν στην γενική κατηγορία του λογισμικού διαχείρισης διαδικασιών (Business process management software - BPMS) είναι η εξής :[22][24],

- *Enterprise Resource Planning (ERP)*. Τα συστήματα ERP χρησιμοποιούνται από τους οργανισμούς για τη διαχείριση επιχειρηματικών διαδικασιών όπως οι προμήθειες, τα logistics, η παραγωγή και οι πωλήσεις.
- *Web data*. Τα δεδομένα που αποθηκεύονται μέσω διαδικτύου και καταγράφουν πληροφορίες σχετικά με τις δραστηριότητες που εκτελούνται από τους χρήστες σε μία ή παραπάνω διαδικτυακές εφαρμογές, όπως επισκέψεις ιστοσελίδων, υποβολή αιτήσεων σε οργανισμού κ.α. Αυτά τα αρχεία καταγραφής μπορούν να χρησιμοποιηθούν για την ανάλυση της συμπεριφοράς των χρηστών και τη βελτιστοποίηση της απόδοσης της εφαρμογής.

<sup>3</sup>Η οργανωσιακή προοπτική εστιάζει σε πληροφορίες σχετικά με πόρους που κρύβονται στο αρχείο καταγραφής, δηλαδή ποιοι παράγοντες (π.χ. άτομα, συστήματα, ρόλοι και τμήματα) εμπλέκονται και πώς σχετίζονται. Ο στόχος είναι είτε η οπτικοποίηση της δομής του οργανισμού ταξινομώντας τους ανθρώπους ως προς τους ρόλους και την οργάνωση σε μονάδες, είτε η εμφάνιση του κοινωνικού δικτύου[21]



- *Customer Relationship Management (CRM)*. Τα συστήματα CRM χρησιμοποιούνται από οργανισμούς για τη διαχείριση των αλληλεπιδράσεων και των δραστηριοτήτων του οργανισμού με τους πελάτες.
- *Operational Databases*. Παράλληλα με τα ERP και CRM, οι εταιρείες μπορεί να χρησιμοποιούν operational database που τα υποστηρίζουν. Οι βάσεις δεδομένων έχουν λειτουργικότητα για την αποθήκευση δεδομένων, ενώ μπορούν να χρησιμεύουν επίσης ως πολύτιμη πηγή δεδομένων που συμπληρώνουν τα αρχεία καταγραφής συμβάντων.
- *Λογισμικό διαχείρισης έργου* Εφαρμογές όπως τα Hive, Trello, ZOHO και JIRA που υποστηρίζουν πολλούς οργανισμούς για την διαχείριση έργων σύμφωνα με συγκεκριμένη μεθοδολογία διεκπεραίωσης (scrum, agile, lean κα.), παράγουν τεράστιο αριθμό δεδομένων σχετικό με τις καθημερινές εργασίες που εκτελούνται.
- *Electronic Health Records (EHRs)* Ψηφιακά αρχεία πληροφοριών για την υγεία των ασθενών. Δημιουργούν αρχεία καταγραφής συμβάντων που μπορούν να χρησιμοποιηθούν για την ανάλυση των αποτελεσμάτων των ασθενών, τη βελτίωση των λειτουργιών του συστήματός υγείας στο οποίο εντάσσονται και τη ποιότητα περίθαλψης.
- *Internet of Things (IoT)* Τα IoT συστήματα αναπτύσσουν μεγάλη δυναμική ως πηγή δεδομένων. Αισθητήρες και ελεγκτές έχουν αναπτυχθεί ευρέως για κάθε είδους σκοπό. Αν και το χάσμα μεταξύ των τυπικών δεδομένων IoT (αναγνώσεις αισθητήρων) και τα αρχεία καταγραφής συμβάντων είναι δύσκολο να γεφυρωθούν, το IoT τείνει να γίνει μια εξαιρετικά σημαντική πηγή εξόρυξης δεδομένων σε επιπλέον τομείς και να δώσει μεγαλύτερες δυνατότητες στην εξόρυξη διαδικασιών καταγράφοντας την ανθρώπινη αλληλεπίδραση με μεγαλύτερη ακρίβεια.

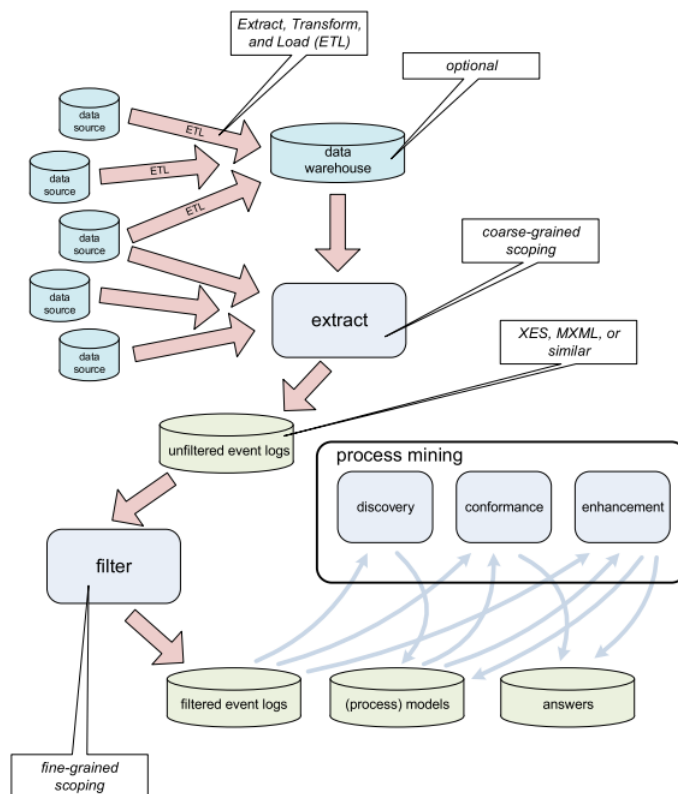


Figure 3.3: Λειτουργία εξόρυξης Διαδικασιών. Πηγή[21]

Τα δεδομένα προκύπτουν από διαφορετικές πηγές και σχεδόν πάντα, ειδικά όταν πρόκειται για μεγάλους οργανισμούς, στην ροή μιας διαδικασίας συμμετέχουν παραπάνω από ένα πληροφοριακά συστήματα. Για παράδειγμα μία προσφορά για παροχή υπηρεσιών μπορεί να πρέπει να αποθηκευτεί ταυτόχρονα σε μια βάση π.χ. Access, να σταλεί στον πελάτη μέσω κάποιου εργαλείου, π.χ. το Jira και να πρέπει να καταχωρηθεί σε ένα ERP π.χ. SAP σε περίπτωση που εγκριθεί για να προχωρήσει η τιμολόγηση της. Αποτέλεσμα είναι η εξόρυξη των δεδομένων και η μετατροπή τους σε αρχεία καταγραφής συμβάντων να είναι ένα δύσκολο και πολύπλοκο έργο που πρέπει να αντιμετωπίσουν οι οργανισμοί. Πρέπει να γίνει κατανοητό ότι η εξόρυξη διαδικασιών βασίζεται σε μεγάλο βαθμό στην ποιότητα των event logs. Κακή ποιότητα δεδομένων μπορεί να οδηγήσει σε λάθος μοντέλα διαδικασιών τα οποία στην συνέχεια θα επηρεάσουν τα αποτελέσματα των αντίστοιχων μεθόδων που θα εφαρμοστούν σε αυτά.

## 3.2 Πρότυπα αρχείων - XES

Οι πολύπλοκες αλληλεπιδράσεις μεταξύ των γεγονότων, των δραστηριοτήτων και των χαρακτηριστικών τους που καταγράφονται κάνει μη λειτουργική την επεξεργασία τους με παραδοσιακούς διδιάστατους πίνακες όπως αρχεία κειμένου και φύλλα Excel[20]. Για τον λόγο αυτό δημιουργήθηκε το πρότυπο XES - eXtensible Event Stream<sup>4</sup> το οποίο είναι μια τυποποιημένη μορφή αναπαράστασης δεδομένων που αφορούν τις δραστηριότητες στο πλαίσιο της εξόρυξης διαδικασιών εγκεκριμένο από τον IEEE

<sup>4</sup><http://www.xes-standard.org/>

Standards Association<sup>5</sup>. Είναι ένα ανοιχτό πρότυπο που χρησιμοποιεί τον ορισμό του σχήματος W3C της γλώσσας XML, καθιερώνει έναν κοινό τρόπο αποθήκευσης, εξαρτήσεων μεταξύ των δραστηριοτήτων και ανταλλαγής δεδομένων επιτρέποντας την επικοινωνία μεταξύ διαφορετικών μεθόδων εξόρυξης δεδομένων.[21].

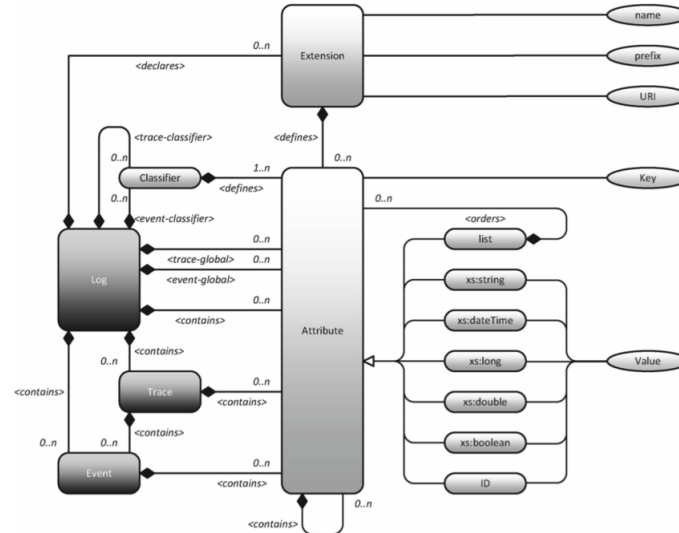


Figure 3.4: Δομή του πρότυπου eXtensible Event Stream (XES). Πηγή [21]

Ένα έγγραφο XES περιέχει μια καταγραφή (log) η οποία αποτελείται από ένα σύνολο από ίχνη (traces)<sup>6</sup>. Κάθε ίχνος περιέχει μια σειρά από γεγονότα, χρονικά ταξινομημένα, που στο σύνολο τους αντιστοιχούν σε μια εκτέλεση περίπτωσης διαδικασίας. Τα τρία κύρια χαρακτηριστικά που αναφέρθηκαν αποτελούνται από ένα σύνολο εμφολευμένων ιδιοτήτων, οι οποίες επιγραμματικά είναι οι *String*, *Date*, *Int*, *Float*, *Boolean*[21]. Είναι ιδανικό για να αντικατοπτρίσει την πραγματική διαδικασία μέσω των μεθόδων εξόρυξης διαδικασιών και το εννοιολογικό πλαίσιο το οποίο αυτή λειτουργεί καθώς μπορεί να εμπλουτιστεί με meta-data, να προστεθούν νέες ιδιότητες, να καθοριστούν κάποιες υποχρεωτικές και να εμφολευτούν.

Συνοψίζοντας, τα XES παρέχουν, αρχικά, μια τυποποιημένη μορφή για την αποθήκευση των event data, διασφαλίζοντας ότι είναι εύκολα διαχειρίσιμα από τα εργαλεία λογισμικού που πραγματοποιούν έργα εξόρυξης διαδικασιών (ProM, Celonis, pm4py κα.). Παρέχουν ευελιξία καθώς μπορεί εύκολα να προστεθούν σε αυτά νέοι τύποι δεδομένων, νέες ιδιότητες και να επεκτείνονται τον "κόσμο" της διαδικασίας. Τέλος, η δυνατότητα να εμπλουτιστεί με meta-data, όπως η προέλευση των δεδομένων, το λογισμικό που χρησιμοποιήθηκε που παρέχουν πρόσθετες πληροφορίες σχετικά με τα αρχεία καταγραφής δεδομένων, κάνουν το XES ένα πραγματικά δυνατό εργαλείο που διαδραματίζει κρίσιμο ρόλο, απαραίτητο στην προώθηση της έρευνας και της πρακτικής [19],[23].

<sup>5</sup><https://standards.ieee.org/about/>

<sup>6</sup>Στον κλάδο της εξόρυξης διαδικασιών ίχνος (trace) θεωρείται μια ακολουθία γεγονότων που σχετίζονται με μια συγκεκριμένη δραστηριότητα. Κάθε συμβάν (event) σε ένα ίχνος αντιπροσωπεύει μια συγκεκριμένη ενέργεια ή εργασία που εκτελέστηκε κατά τη διάρκεια ενός Process Instance. Ένα ίχνος μπορεί να θεωρηθεί ως μια χρονολογική καταγραφή των γεγονότων που συνέβησαν κατά τη διάρκεια της εκτέλεσης της διαδικασίας και παρέχει μια λεπτομερή εικόνα του τρόπου με τον οποίο εκτελέστηκε πραγματικά [19].

### 3.3 Μοντέλα διαδικασιών

Τα μοντέλα διαδικασιών χρησιμοποιούνται για να παρουσιάσουν, να αναλύσουν και πολλές φορές να προβλέψουν την συμπεριφορά μιας διαδικασίας. Η χρήση τους είναι απαραίτητη από τους οργανισμούς καθώς δίνουν σαφείς οδηγίες και εφαρμόζουν εταιρικούς κανόνες. Καθορίζουν το σχετικό πλαίσιο εργασίας που χρειάζονται οι διαδικασίες ενός οργανισμού που τείνουν να γίνουν πολύπλοκες εξαιτίας της αυξανόμενης αλληλεπίδρασης πληροφοριακών συστημάτων και μελών ενός οργανισμού και την συμμόρφωσή τους με την επιχειρησιακή πολιτική και τους κρατικούς νόμους. Η σημασία και η ανάγκη της μοντελοποίησης των διαδικασιών αποδεικνύεται από την πληθώρα των διαγραμμάτων, εφαρμογών και εργαλείων που χρησιμοποιούνται για την κατασκευή τους.

Τεχνικά πρόκειται για γραφικές αναπαραστάσεις όπως χάρτες διαδικασιών και διαγράμματα ροής που απεικονίζουν μια ακολουθία των δραστηριοτήτων, των αποφάσεων και των αλληλεπιδράσεων που συμμετέχουν σε μια συγκεκριμένη επιχειρησιακή διαδικασία. Αποτελεί βασικό συστατικό ενός έργου εξόρυξης διαδικασιών, ώστε να παρέχει μια οπτική αναπαράσταση των δραστηριοτήτων και των εργασιών που συμμετέχουν σε μια διαδικασία με σκοπό την περαιτέρω ανάλυση, την βελτιστοποίηση και την βελτίωση της διαδικασίας. Κατανοεί τις σχέσεις μεταξύ των διαφόρων δραστηριοτήτων, την αναγνώριση πιθανών προβλημάτων και αδυναμιών, καθώς και την αναζήτηση τρόπων βελτίωσης και αποτελεσματικότερης ροής των διαδικασιών. Επιπλέον, τα μοντέλα διαδικασιών μπορούν να χρησιμοποιηθούν ως εργαλεία επικοινωνίας, επιτρέποντας στους εμπλεκόμενους να κατανοήσουν τον τρόπο λειτουργίας μιας διαδικασίας και να συμβάλουν στην ανάπτυξη και βελτίωσή της.

Η δημιουργία μοντέλων είναι μια δύσκολη, απαιτητική και επιρρεπή σε λάθη εργασία εξαιτίας της ιδεατής λειτουργίας της διαδικασίας που έχει ο κατασκευαστής της και δεν μπορεί να αντιληφθεί τα λάθη, η ανθρωπινή συμπεριφορά η οποία είναι απρόβλεπτη και μπορεί να δημιουργήσει μεγάλες αποκλίσεις στην διαδικασία και το μοντέλο να έχει εσφαλμένο επίπεδο αφάρεσης και λάθος είσοδο δεδομένων ή εξαιτίας των λάθος ερωτήσεων που πρέπει να απαντηθούν μέσω του μοντέλου. Η επιστήμη της εξόρυξης δεδομένων αντίθετα παρέχει την δυνατότητα κατασκευής, σχετικά εύκολα, μιας ποικιλίας διαγραμμάτων και μοντέλων βασισμένα σε αρχεία καταγραφής γεγονότων τα οποία ανακαλύπτουν την διαδικασία όπως πραγματικά είναι. Παρακάτω παρουσιάζεται μια επισκόπηση των βασικών εννοιών, τεχνικών και εργαλείων που χρησιμοποιούνται για την αναπαράσταση διαδικασιών.

#### 3.3.1 PetriNet

Τα διαγράμματα PetriNet είναι η θεμελιώδης σημειογραφία για την οπτικοποίηση των διαδικασιών. Πρόκειται για ένα γραφικό μοντέλο που αναπαριστά ένα στατικό δίκτυο το οποίο δεν αλλάζει κατάσταση από μόνο του και αναπαριστά την αλληλεπίδραση των οντοτήτων σε ένα σύστημα διαδικασίας. Αποτελεί ένα ισχυρό εργαλείο για την παρουσίαση και ανάλυση της συμπεριφοράς πολύπλοκων συστημάτων. Τα βασικά στοιχεία που αποτελείται ένα PetriNet είναι οι καταστάσεις (places), οι μεταβάσεις (transitions), οι ενδείξεις (tokens) και τα βέλη (arcs).

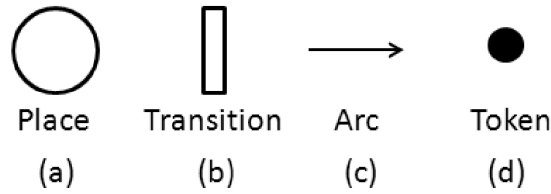


Figure 3.5: Οντότητες ενός PetriNet. Πηγή [3]

Τα tokens έχουν την ιδιότητα να μετακινούνται από την μία κατάσταση στην άλλη (place to place) και η κάθε κατάσταση αποθηκεύει ένα token. Οι μεταβάσεις παράγουν ή καταναλώνουν tokens και τα arcs ενώνουν τις μεταβάσεις με τουλάχιστον μία κατάσταση και αντίθετα. Η σημειογραφία αλλάζει μορφή όταν πυροδοτηθεί μία μετάβαση, και κάθε μετάβαση που πυροδοτείται καταναλώνει ένα token από κάθε μία κατάσταση που έχει σαν είσοδο και αντίστοιχα να παράγει τόσα token όσα τα arcs που οδηγούν στις αμέσως επόμενες καταστάσεις.

Οι βασικές ιδιότητες ενός PetriNet είναι οι boundness, deadlocks και liveness οι οποίες επιτρέπουν μια λεπτομερή ανάλυση του μοντελοποιημένου συστήματος. Εν συντομία:

- *Boundness*, αναφέρεται στην ιδιότητα του PetriNet να μπορεί να έχει σαν όριο έναν μέγιστο αριθμό token σε κάθε κατάσταση. Μία κατάσταση 'p' είναι *k*-bounded αν δεν είναι προσβάσιμη σε περισσότερα από 'k' tokens στην κατάσταση 'p'. Ένα PetriNet είναι boundness, εάν ο αριθμός των tokens σε κάθε κατάσταση δεν υπερβαίνει τον ανώτατο αριθμό 'k' για οποιαδήποτε κατάσταση προσβάσιμη από το μοντέλο.
- *Deadlock*, μια κατάσταση ορίζεται σαν "dead" αν καμία μετάβαση δεν την ενεργοποιεί. Ένα PetriNet λέγεται ότι έχει deadlock αν υπάρχει μια κατάσταση 'dead'. Για να χαρακτηριστεί ένα PetriNet "deadlock free" πρέπει να επιτρέπει για οποιαδήποτε κατάσταση τουλάχιστον μια μετάβαση.
- *Liveness*, μια μετάβαση 't' είναι "live" αν και μόνο αν από μια προσβάσιμη ως προς αυτήν κατάσταση φτάσουμε σε οποιαδήποτε άλλη κατάστασή μέσα στο μοντέλο πυροδοτώντας την μετάβαση 't'. Το μοντέλο είναι "live" αν όλες οι μεταβάσεις του είναι "live". Αν ένα μοντέλο είναι "live" τότε είναι και μοντέλο χωρίς αδιέξοδα (deadlock free).

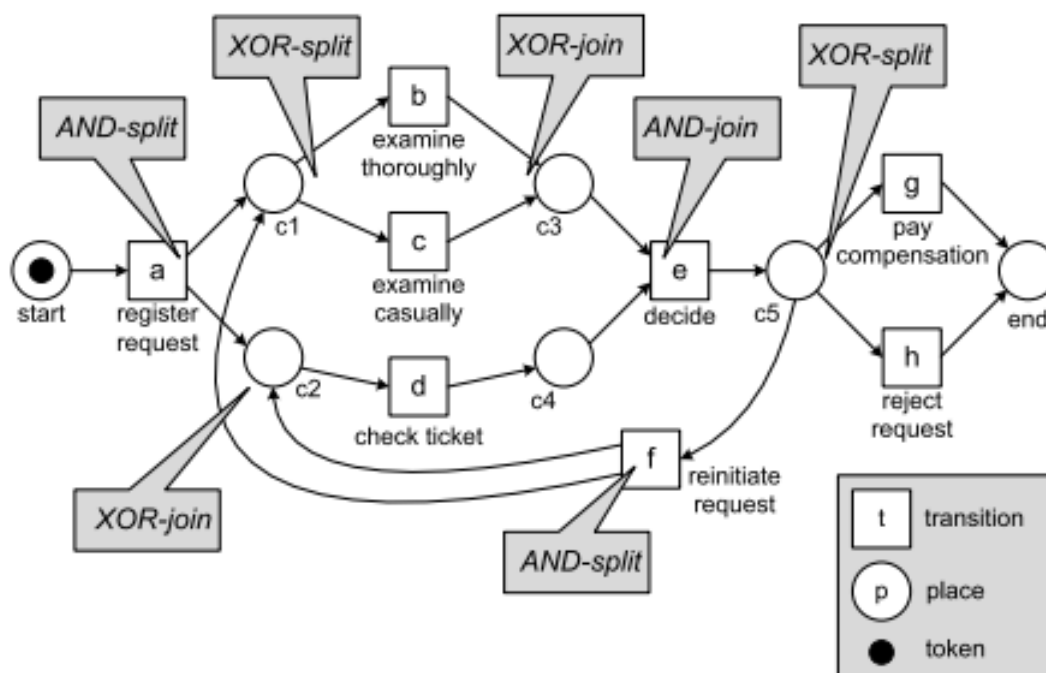


Figure 3.6: Παράδειγμα ενός PetriNet. Πηγή [21]

Μια κατηγορία των διαγραμμάτων PetriNets είναι τα WorkFlow-Nets στα οποία γίνεται αναφορά γιατί είναι ο στόχος του αλγόριθμου 'A' είναι να εξάγει ένα τέτοιου είδους διάγραμμα. Ένα τέτοιο δίκτυο έχει μία αρχική θέση, μια τελική θέση και όλες οι μεταβάσεις και οι καταστάσεις να βρίσκονται σε μια θέση από την αρχή μέχρι το τέλος. Για να χαρακτηριστεί ένα PetriNet σαν WorkFlow-Net πρέπει να έχει συγκεκριμένες ιδιότητες. Αρχικά πρέπει να είναι *safe*, δηλαδή οι καταστάσεις(places) δεν πρέπει να κρατάνε πολλά tokens την ίδια στιγμή, η διαδικασία να ολοκληρώνεται πάντα και όταν φτάσει στην τελική κατάσταση δεν πρέπει να έχουν ξεμείνει tokens μέσα στο δίκτυο και τέλος να μην υπάρχουν *dead* καταστάσεις, πρέπει όλες να είναι ενεργοποιημένες.

### 3.3.2 BPMN

Το διάγραμμα BPMN (Business Process Modeling Notation) συστήθηκε το 2004 και έχει καθιερωθεί ως η πιο συχνά χρησιμοποιούμενη τεχνική για την αναπαράσταση επιχειρησιακών διαδικασιών και αποτελεί μια ισχυρή εργαλειοθήκη για τη μοντελοποίηση και τη διαχείριση τους. Απεικονίζουν την ακολουθία των δραστηριοτήτων, των γεγονότων, των αποφάσεων και την ροή δεδομένων πληροφοριών μέσα στην διαδικασία. Προδιαγράφεται από το OMG Standard <sup>7</sup> και υποστηρίζεται από πολλούς κατασκευαστές λογισμικού γιατί παρέχει μια κοινή γλώσσα για την επικοινωνία και την κατανόηση των πολύπλοκων διαδικασιών που υπάρχουν στους οργανισμούς, είναι κοντά στο UML διάγραμμα δραστηριοτήτων και τα Event-Driven Process chains.

<sup>7</sup><https://www.omg.org/about/omg-standards-introduction.htm>

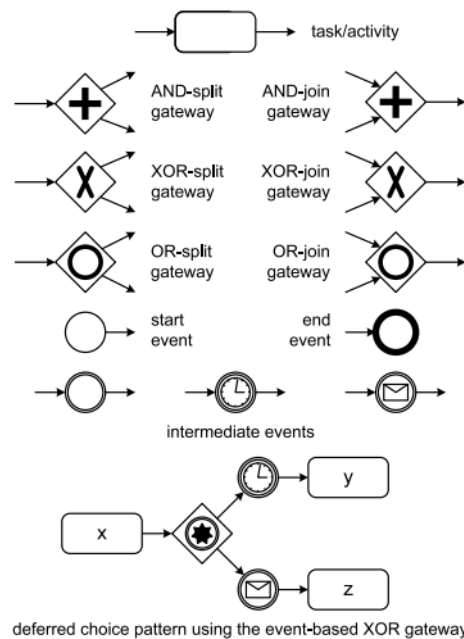


Figure 3.7: Σημειογραφία BPMN. Πηγή [21]

Τα διαγράμματα BPMN αποτελούνται από συγκεκριμένες οντότητες οι οποίες απεικονίζουν την διαδικασία. Συγκεκριμένα, αποτελούνται από τις δραστηριότητες (Task, Process, Sub-process) που αναπαριστούν συγκεκριμένες εργασίες μέσα σε μια διαδικασία και κατηγοριοποιούνται ως απλές εργασίες, υπόδιαδικασίες ή συναλλαγές. Τα γεγονότα (events) τα οποία απεικονίζουν μία ενέργεια κατά την διάρκεια της διαδικασίας, όπως τα σημεία εκκίνησης, ολοκλήρωσης, κ.α. τα οποία μπορούν να πυροδοτηθούν από διάφορους παράγοντες όπως η λήψη ενός μηνύματος ηλεκτρονικού ταχυδρομείου. Τις πύλες (Gateways) οι οποίες απεικονίζουν σημεία απόφασης και διακλάδωσης μέσα σε μια διαδικασία καθορίζοντάς την ροή της βάσει κανόνων και συνθηκών. Τις ροές (Sequence/Message Flows) οι οποίες εμφανίζουν τις συνδέσεις μεταξύ των δραστηριοτήτων, των γεγονότων και των πυλών που αναπαριστούν την ακολουθία εκτέλεσης και την κατεύθυνση της διαδικασίας. Τέλος μπορεί να προστεθούν επιπλέον στοιχεία τεκμηρίωσης τα οποία παρέχουν επιπλέον πληροφορίες που βοηθούν στην καλύτερη αναπαράσταση της διαδικασίας.

Τα οφέλη του BPMN είναι αρκετά και καθιστούν την χρήση τους διαδεδομένη. Αρχικά παρέχει ολοκληρωμένες και πλούσιες σημειογραφίες που μπορούν εύκολα να κατανοηθούν ώστε να διευκολυνθεί η σαφής επικοινωνία και κατανόηση της διαδικασίας μεταξύ των εμπλεκόμενων οργανισμών ανεξάρτητα από το τεχνολογικό και επιχειρηματικό υπόβαθρο των μελών τους. Παρέχει στις επιχειρήσεις μέσω των οπτικών αναπαραστάσεων τη δυνατότητα να ορίζουν και να κατανοούν τις διαδικασίες μέσω των διαγραμμάτων και επιτρέπει στους χρήστες να αντιλαμβάνονται την δομή τους. Τέλος, το BPMN σε συνδυασμό με την εξόρυξη διαδικασιών επιτρέπει την ανάλυση και αναγνώριση προβληματικών περιοχών, των μοτίβων με αποτέλεσμα την βελτιστοποίηση και αναδιοργάνωση της διαδικασίας.

### 3.3.3 Directly-Follows Graph

Ο Directly-Follow Graph (DFG) ενός αρχείου καταγραφής συμβάντων είναι ένας κατευθυνόμενος γράφος όπου κάθε κόμβος αντιπροσωπεύει μια δραστηριότητα που αναφέρεται στο αρχείο καταγραφής συμβάντων, και κάθε ακμή αντιπροσωπεύει μια σχέση άμεσης ακολουθίας. Για να καθορίσουμε ξεκάθαρα την αρχή και το τέλος της διαδικασίας, συνηθίζεται να περιλαμβάνεται ένας ξεκάθαρος κόμβος αρχής και ένας ξεκάθαρος κόμβος τέλους στο DFG ο οποίος είναι αποτέλεσμα ενός του Inductive αλγόριθμου εξόρυξης διαδικασιών[4]. Τα DFGs, επειδή είναι αρκετά κατανοητά από τους χρήστες ανεξαρτήτως επαγγελματικού υπόβαθρου, χρησιμοποιούνται σε ευρεία χρήση ως αυτόνομη γραφική αναπαράσταση στα εμπορικά εργαλεία εξόρυξης διαδικασιών.

Ο DFG οποίος προέρχεται από παραγωγικά αρχεία καταγραφής, τα οποία συνήθως είναι μεγάλα σε όγκο, συνήθως περιέχει πολλούς κόμβους και μεγάλο πλήθος ακμών κάτι που δυσχεραίνει την κατανόηση τους. Για να αντιμετωπιστεί η πολυπλοκότητα, οι DFGs απλοποιούνται αρκετά εύκολα με την αφαίρεση κόμβων και ακμών με βάση την συχνότητα. Οι αναλυτές και οι τεχνικοί που υλοποιούν ένα έργο εξόρυξης διαδικασιών τείνουν να χρησιμοποιούν ενεργά τέτοιους γράφους, οι οποίοι παρά τη απλότητά τους, μπορεί να τους παραπλανήσουν για αυτό και πρέπει να γνωρίζουν πώς δημιουργούνται αυτά τα μοντέλα διαδικασίας πριν τα ερμηνεύσουν. Οι απλοποιημένοι DFG πρέπει να διατηρούν δύο ιδιότητες, η πρώτη είναι ότι πρέπει να διατηρείται το σύνολο των κόμβων του αρχικού-πλήρους γράφου και η δεύτερη ιδιότητα είναι ότι κάθε κόμβος στο φιλτραρισμένο DFG πρέπει να βρίσκεται σε μια διαδρομή από τον κόμβο που αναπαριστά την έναρξη της διαδικασίας έως τον κόμβο που αναπαριστά την τερματική δραστηριότητα της.

### 3.3.4 Δέντρα διαδικασιών

Τα δέντρα διαδικασιών (Process Tree) είναι γραφικές αναπαραστάσεις που έχουν σχεδιαστεί ειδικά για την εξόρυξη των διαδικασιών με στόχο να αποτυπώσουν την παρατηρούμενη συμπεριφορά της διαδικασίας με έναν δομημένο και κατανοητό ως προς τον χρήστη τρόπο. Είναι μια ιεραρχική δομή όπου οι εσωτερικοί κόμβοι είναι οι τελεστές (sequential composition, exclusive choice, parallel composition και redo loop) και τα φύλλα οι δραστηριότητες της διαδικασίας[21].

Τα κοινά διαγράμματα που αναφέρθηκαν όπως τα PetriNet και τα BPMN έχουν διάφορα μειονεκτήματα όπως αδιέξοδα και ανεπαρκείς διαδρομές (unsound) με αποτέλεσμα να μην μπορεί να περιγράψει επακριβώς την παρατηρούμενη συμπεριφορά. Αντίθετα τα δέντρα διαδικασίας είναι από κατασκευής αξιόπιστα εξαιτίας της χρήσης των inductive τεχνικών ανακάλυψης του μοντέλου διαδικασίας οι οποίες εξασφαλίζουν την ακεραιότητα και το γεγονός ότι μπορούν να διαχειριστούν σπάνιες συμπεριφορές και μεγάλα σε μέγεθος μοντέλα και αρχεία καταγραφής γεγονότων. Επίσης, τα δέντρα διαδικασίας παρέχουν μια καθαρή και ευανάγνωστη απεικόνιση της ροής της διαδικασίας, καθιστώντας τα προσίτα σε χρήστες με τεχνικό ή μη υπόβαθρο γεγονός που διευκολύνει την αποτελεσματική επικοινωνία και κατανόηση της διαδικασίας μεταξύ των εμπλεκόμενων.



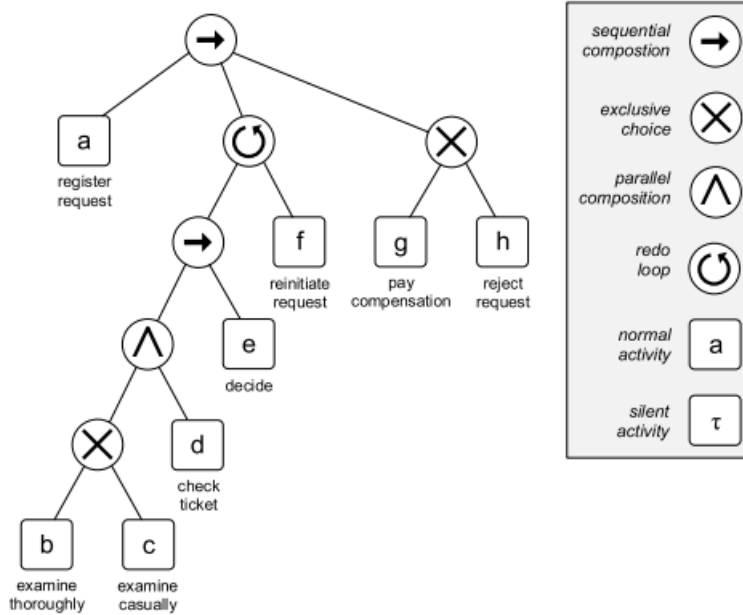


Figure 3.8: Παράδειγμα ενός δένδρου διαδικασιών και η σημειογραφία των τελεστών του. Πηγή [21]

### 3.4 Αλγόριθμοι Ανακάλυψης Διαδικασιών

Όπως έχει αναφερθεί αρχικό στάδιο ενός έργου εξόρυξης διαδικασιών είναι η ανακάλυψη της διαδικασίας, πρόκειται για μία ανάλυση που βασίζεται στον έλεγχο και την ανακάλυψη της ακολουθίας των δραστηριοτήτων. Έτσι, σαν στόχο έχει την καταγραφή της ακολουθίας των δραστηριοτήτων που βρίσκονται στο αρχείο καταγραφής συμβάντων και την δημιουργία του βέλτιστου μοντέλου διαδικασιών. Για αυτόν το λόγο χρησιμοποιούνται αρκετοί αλγόριθμοι ανακάλυψης διαδικασιών όπως ο a-miner, ο heuristic miner, ο genetic miner, ο inductive miner κ.α. που έχουν σαν αποτέλεσμα, ανεξαρτήτως της προσέγγισης που καθένας χρησιμοποιεί, να δημιουργείται ένα μοντέλο το οποίο απεικονίζει τις μεταβάσεις μεταξύ των δραστηριοτήτων.

Βασική ιδέα όλων το αλγορίθμων είναι, κατά την διάρκειά ανάγνωσης του αρχείου καταγραφής συμβάντων, αν μια δραστηριότητα  $a$  ακολουθείται από μια δραστηριότητα  $b$ , τότε η μετάβασή αυτή καταμετράται. Στην συνέχεια επαναλαμβάνεται η ίδια διαδικασία για όλα τα ζεύγη συνεχόμενων εργασιών που έχουν το ίδιο αναγνωριστικό-id. Με αυτόν τον τρόπο υπάρχει η πληροφορία για το πόσες φορές έχει συμβεί μια συγκεκριμένη μετάβαση. Τέλος, γίνεται ο συνδυασμός αυτών των μεταβάσεων οι μεταξύ τους σχέσεις ώστε να δημιουργηθεί ένας γράφος που θα απεικονίζει την συμπεριφορά της διαδικασίας.

Οι προκλήσεις που αντιμετωπίζει ένας αλγόριθμος εξόρυξης διαδικασιών είναι, αρχικά, η κλάση αναπαράστασης του μοντέλου (representational bias), η οποία είναι ουσιαστικά ο τρόπος με τον οποίο κατασκευάζεται και αναπαριστάται το μοντέλο διαδικασίας. Είναι σημαντικό χαρακτηριστικό γιατί επηρεάζει τα αποτελέσματα της ανάλυσης. Παραδείγματά κλάσεων των μοντέλων διαδικασιών είναι τα PetriNets και τα διαγράμματα BPMN και ανάλογα με πιο χρησιμοποιείται μπορεί να τονιστούν διαφορετικές πτυχές της διαδικασίας και μπορεί τα αποτελέσματα να είναι λιγότερο ή περισσότερο χρήσιμα ανάλογα συγκεκριμένα χαρακτηριστικά της διαδικασίας που

αναλύεται. Ο κάθε αλγόριθμος έχει καθορισμένο είδος αναπαράστασης που μπορεί να παράξει, το οποίο καθορίζει το χώρο αναζήτησης και περιορίζει την εκφραστικότητά του μοντέλου που ανακαλύφθηκε. Ο αλγόριθμος-α κατασκευάζει ένα μοντέλο PetriNet χωρίς να λαμβάνει υπόψιν του επαναλαμβανόμενες δραστηριότητες, και παρουσιάζει τις δραστηριότητες ως κόμβους και τις μεταβάσεις σαν ακμές. Ο Heuristic Miner επίσης μπορεί να παράξει PetriNets και C-Nets(Casual Nets) χρησιμοποιώντας την συχνότητα εμφάνισης δραστηριοτήτων η οποία ενδέχεται να μην καταγράψει μια σπάνια συμπεριφορά, την σειρά με την οποία εκτελούνται οι δραστηριότητες και τις μεταξύ τους εξαρτήσεις. Τέλος ο αλγόριθμος Inductive Miner χρησιμοποιεί δέντρα διαδικασιών, PetriNets, BPMN για να αποτυπώσει τις εξαρτήσεις δεδομένων μεταξύ των δραστηριοτήτων, χρησιμοποιώντας την επαγωγική σκέψη. Η κλάση του μοντέλου επικεντρώνεται στις δραστηριότητες και όχι στα δεδομένα που αλληλεπιδρούν και για αυτό τον λόγο μπορεί να οδηγήσει σε ανεπαρκή μοντέλα διεργασιών που δεν καταγράφουν πλήρως τις εξαρτήσεις δεδομένων στη διαδικασία. Το representational bias είναι σημαντικό χαρακτηριστικό γιατί μπορεί να έχει σημαντικές επιπτώσεις στην ορθότητα και τη χρησιμότητα των αποτελεσμάτων της διαδικασίας εξόρυξης διαδικασιών. Εάν η αναπαράσταση του μοντέλου διαδικασίας ή ο αλγόριθμος εξόρυξης διαδικασιών δεν είναι κατάλληλα προσαρμοσμένος στα ειδικά χαρακτηριστικά της διαδικασίας που αναλύεται, τα αποτελέσματα μπορεί να είναι παραπλανητικά ή ανεπαρκή.

Επιπλέον, ένας αλγόριθμος εξόρυξης διαδικασιών πρέπει να είναι σε θέση να διαχειρίζεται τον θόρυβο(noise), δηλαδή συμπεριφορές που εμφανίζονται σπάνια ή αποτελούν εξαιρέσεις, και την μη πληρότητα(incompleteness) των αρχείων καταγραφής, δηλαδή ένα μικρό σε μέγεθος σύνολο γεγονότων, μη επαρκές για την εξαγωγή συμπερασμάτων ή ανακάλυψη συγκεκριμένης συμπεριφοράς[21]. Συγκεκριμένα ο θόρυβος οφείλεται στην παρουσία μη σχετικών, εσφαλμένων ή παραπλανητικών πληροφοριών στο αρχείο γεγονότων και μπορεί να είναι αποτέλεσμα ανθρώπινου σφάλματος, σφάλματος στο σύστημα ή προβλήματα συλλογής δεδομένων και να επηρεάζει την ακρίβεια και αξιοπιστία των αποτελεσμάτων της διαδικασίας ανακάλυψης διαδικασιών. Η μη πληρότητα οφείλεται στην έλλειψη πληροφοριών ή στην απουσία δεδομένων στο αρχείο καταγραφής γεγονότων. Συμβαίνει όταν δεν καταγράφονται ορισμένα γεγονότα ή ακολουθίες δραστηριοτήτων, χαρακτηριστικά που σχετίζονται μεταξύ τους και δεδομένα που αφορούν τις δραστηριότητες. Είναι αποτέλεσμα τεχνικών περιορισμών, μη καταγραφής σημαντικών δεδομένων εξαιτίας ανησυχιών για την προστασία των δεδομένων και επιλεκτική καταγραφή.

Τέλος, το μοντέλο που παράγει ένας αλγόριθμος εξόρυξης διαδικασιών αξιολογείται βάσει συγκεκριμένων κριτηρίων. Τα κριτήρια ποιότητάς με τα οποία γίνεται η αξιολόγηση ενός μοντέλου διαδικασίας, στην βιβλιογραφία αναφέρονται σαν διαστάσεις. Είναι η καταλληλότητα(fitness), η απλότητα(simplicity), η ακρίβεια(precision) και η γενίκευση(generalization)[21].

Η καταλληλότητα(fitness) αναφέρεται στον βαθμό τον οποίο ένα μοντέλο διαδικασίας αναπαριστά ακριβώς, αναφορικά με την συμπεριφορά που έχει καταγραφεί στο αρχείο καταγραφής γεγονότων. Αν ένα μοντέλο έχει υψηλό ποσοστό καταλληλότητάς τότε αντιστοιχεί στις πραγματικές εκτελέσεις δραστηριοτήτων και την ροή της διαδικασίας στον πραγματικό κόσμο. Ο βαθμός της απλότητας(simplicity) σχετίζεται με το πόσο κατανοητό, το πόσο σωστά κατανεμημένο είναι ένα μοντέλο διαδικασίας και το πόσο εύκολο είναι να ερμηνευτεί από τους αναλυτές. Αντίθετα η ακρίβεια (precision) αναφέρεται στο επίπεδο λεπτομέρειας και ακρίβειας των αποτελεσμάτων ανάλυσης. Ακριβή αποτελέσματα εξασφαλίζουν ότι μοντέλο διαδικασίας δεν περιέχει ασάφειες και

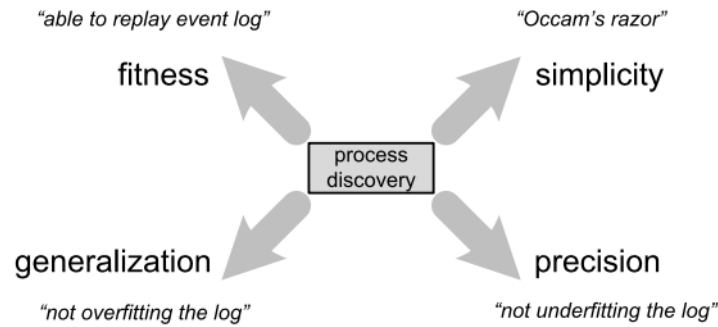


Figure 3.9: Εξισορρόπηση των τεσσάρων κριτηρίων ποιότητας: καταλληλότητα, απλότητα, ακρίβεια και γενίκευση. . Πηγή [21]

είναι απαλλαγμένο από σφάλματα. Τέλος, η γενίκευση (generalization) σχετίζεται με την ικανότητα ενός μοντέλου να αποτυπώσει την κοινή συμπεριφορά και τα υποκείμενα μοτίβα σε ένα εύρος περιπτώσεων διαδικασίας. Ένα γενικευμένο μοντέλο μπορεί να παρέχει περιπτώσεις που είναι εφαρμόσιμες πέρα από αυτές που χρησιμοποιούνται για την ανάλυση.

Table 3.1: Περιληπτική απεικόνιση με τα πλεονεκτήματα και τα μειονεκτήματα των αλγορίθμων εξόρυξης διαδικασιών. Πηγή [16].

Alpha Miner	Alpha Miner+	Heuristic	Inductive
Δεν μπορεί να διαχειριστεί βρόγχους μήκους 1 και 2.	Μπορεί να διαχειριστεί βρόγχους μήκους 1 και 2.	Υπολογίζει το κατά πόσο συχνά πραγματοποιείται η διαδικασία (frequency).	Μπορεί να χειριστεί αόρατες εργασίες.
Οι αόρατες και διπλά εισαγόμενες εργασίες δεν μπορούν να ανακαλυφθούν.	Οι αόρατες και διπλά εισαγόμενες εργασίες δεν μπορούν να ανακαλυφθούν.	Ανιχνεύει σύντομους βρόγχους.	Το μοντέλο είναι ορθό.
Το ανακαλυφθέν μοντέλο ενδέχεται να μην είναι ακριβές (sound)	Το ανακαλυφθέν μοντέλο ενδέχεται να μην είναι ακριβές (sound)	Δεν εγγυάται ότι θα ανακαλυφθεί ένα ακριβές μοντέλο (sound)	Ο πιο συχνά χρησιμοποιούμενος αλγόριθμος εξόρυξης διαδικασιών
Αδύναμο σε logs που περιέχουν θόρυβο	Αδύναμο σε logs που περιέχουν θόρυβο		

### 3.4.1 Alpha Miner

Ο Alpha Miner είναι πρώτος αλγόριθμος εξόρυξης διαδικασιών που παρουσιάστηκε από τους Aalst et al. το 2004 [20]. Ο πρωταρχικός του στόχος είναι να κατασκευάσει ένα PetriNet χρησιμοποιώντας με βάση την παρατηρούμενη συμπεριφορά όπως αυτή έχει αποθηκευτεί στα αρχεία καταγραφής γεγονότων. Ο αλγόριθμος εστιάζει στην καταγραφή των σχέσεων μεταξύ των δραστηριοτήτων σε μια διαδικασία και την παρουσιάζει με μορφή ακολουθίας. Ήταν ο πρώτος ο οποίος μπορούσε να διαχειριστεί τις διαδικασίες που εκτελούνται ταυτόχρονα. Σε γενικές γραμμές ο αλγόριθμος-α αποτελεί μια καλή βάση για να κατανοηθούν οι προκλήσεις που αντιμετωπίζονται στην εξόρυξη διαδικασιών λόγω τις απλότητας του και το γεγονός ότι πολλές ιδέες του

εφαρμόζονται σε πολύπλοκες τεχνικές. Έχει ωστόσο αρκετούς περιορισμούς καθώς δεν μπορεί να διαχειριστεί τον θόρυβο, τις σπάνιες περιπτώσεις, τις ημιτελείς και εκείνες που ακολουθούν πολύπλοκες διαδρομές μέσα στο διάγραμμα.

Σαν είσοδο έχει ένα αρχείο καταγραφής γεγονότων, οι δραστηριότητες των οποίων μεταφράζονται στις μεταβάσεις ενός PetriNet και σαν έξοδο ένα γράφημα PetriNet με ιδανικό στόχο ένα WorkFlowNet. Ο αλγόριθμος- $a$  σαρώνει το αρχείο και ψάχνει για συγκεκριμένα μοτίβα. Για παράδειγμα, αν μια δραστηριότητα  $a$  ακολουθείται από μία  $\beta$ , αλλά η  $\beta$  δεν ακολουθείται ποτέ από την  $a$ , τότε υποθέτουμε ότι υπάρχει αιτιώδης σχέση μεταξύ των  $a$  και  $\beta$ . Για να αντικατοπτρίσει αυτήν την εξάρτηση, το αντίστοιχο PetriNet πρέπει να έχει έναν χώρο που συνδέει το  $a$  με το  $\beta$ . [21]

Άρα το πρώτο βήμα του αλγόριθμου είναι να είναι να ληφθούν υπόψη οι σχέσεις τάξης. Για αυτό τον λόγο διακρίνουμε τέσσερις σχέσεις τάξης. Συγκεκριμένα για ένα σύνολο  $L$  που αντιστοιχεί σε ένα αρχείο καταγραφής που περιέχει ένα σύνολο δραστηριοτήτων 'A' ισχύει [21],

- $a >_L b$ , εάν και μόνο εάν υπάρχει μία περίπτωση η οποία  $\sigma = t_1, t_2, t_3, \dots, t_n$  και  $i \in 1, \dots, n - 1$
- $a \rightarrow_L b$  εάν και μόνο εάν  $a >_L b$  και  $b \not>_L a$
- $a \# b$  εάν και μόνο εάν  $a \not>_L b$  και  $b \not>_L a$
- $a \parallel b$  εάν και μόνο εάν  $a >_L b$  και  $b >_L a$ .

Επόμενο βήμα είναι βάσει αυτών το σχέσεων να ανακαλυφθούν συγκεκριμένα μοτίβα συμπεριφοράς και χρησιμοποιώντας τα γραφήματα PetriNet να δημιουργήσει το αντίστοιχο μοντέλο. Αναδεικνύονται, όπως αναφέρθηκε, σημαντικοί περιορισμοί. Αρχικά υποθέτει ότι το αρχείο καταγραφής συμβάντων είναι πλήρες, πράγμα που σημαίνει ότι για να ανακαλυφθεί ένα σωστό μοντέλο πρέπει να έχετε δει όλα τα πιθανά ίχνη της διαδικασίας μας, στην παραγωγική αυτό δεν είναι ρεαλιστικό. Επιπλέον δεν μπορεί να διαχειριστεί τα 'self-loop' μιας δραστηριότητας, δηλαδή να επαναλαμβάνεται συνεχόμενα, γιατί αν μια δραστηριότητα  $\beta$  ακολουθείται από την ίδια δραστηριότητα  $\beta$  και κάποιες φορές το αντίστροφο, στην καλύτερη περίπτωση αυτό θα αποτυπωθεί σαν παράλληλη σχέση και ο αλγόριθμος- $a$  θα βάλει αυτή τη δραστηριότητα σαν μια μετάβαση, άρα θα αποτυπώσει μια διαφορετική διαδικασία στο μοντέλο. Τέλος, παρόμοιο πρόβλημα αντιμετωπίζεται και με τα 'short-loops' που περιέχουν λίγες δραστηριότητες (δύο με τρεις) που εκτελούνται επαναληπτικά, γιατί, επίσης, δεν υπάρχει αντίστοιχος κανόνας που να μπορεί να τις ανακαλύψει και θα κάνει υπόθεση ότι αυτές οι διαδικασίες εκτελούνται παράλληλα.

### 3.4.2 Inductive Miner

Σαν προσέγγιση η χρήση του Inductive αλγόριθμου με σκοπό την ανακάλυψη μιας διαδικασίας είναι από τις κορυφαίες λόγω της ευελιξίας της, της επιβεβαίωσης και της επεκτασιμότητας που προσφέρει. Υπάρχουν αρκετές τεχνικές που όλες όμως χρησιμοποιούν τα δέντρα διαδικασίας (Process Trees) όπου σε αντίθεση με άλλες σημειογραφίες όπως PetriNet, BPMN, Casual Nets, είναι από κατασκευής ακέραια (sound) και δεν αντιμετωπίζουν τα προβλήματα των υπολοίπων. Επίσης, όπως έχει αναφερθεί, ένα δέντρο διαδικασίας μετατρέπεται εύκολα σε οποιαδήποτε άλλο γράφημα δίνοντάς ένα ιδανικό

μοντέλο διαδικασιών ως προς την παρατηρούμενη συμπεριφορά. Ο συγκεκριμένος αλγόριθμος μπορεί να ανακαλύψει ένα μεγαλύτερο φάσμα διαδικασιών και να εξορύξει τα σωστά μοντέλα στα σημεία όπου αλγόριθμος- $a$  και οι υπόλοιποι αποτυγχάνουν.

Όπως αναφέρθηκε σκοπός του "Inductive Miner", είναι να ανακαλύψει ένα δένδρο διαδικασιών που θα αναπαριστά την διαδικασία. Δεδομένου ενός αρχείου καταγραφής συμβάντων, κατασκευάζεται ένας "directly-follows graph" χρησιμοποιώντας τις αντίστοιχες ιδιότητές του (την σχέση "directly-follows" ( $>_L$ ), ένα σύνολο δραστηριοτήτων εκκίνησης και ένα σύνολο δραστηριοτήτων τερματισμού). Σαν σκοπό έχει να ανακαλύψει στο αρχείο καταγραφής μοτίβα και δραστηριότητες που έχουν να κάνουν κυρίως με αποφάσεις, ακολουθίες, δραστηριότητες που εκτελούνται παράλληλα ή επαναλαμβανόμενα, να εφαρμόσει τους αντίστοιχους τελεστές και να διαχωρίσει το αρχείο καταγραφής σε μικρότερα αρχεία μέχρι να επιτευχθεί μια βασική δραστηριότητα, δηλαδή ένα "event log" με μόνο μία δραστηριότητα.

Ο Inductive Miner πλεονεκτεί έναντι των άλλων αλγορίθμων γιατί μπορεί να διαχειρίζεται θορυβώδη και ημιτελή δεδομένα που έχουν καταγραφεί στο event log, καθιστώντας τον πλέον κατάλληλο για παραγωγικές και πολύπλοκες διαδικασίες. Δημιουργεί μοντέλα διαδικασιών που είναι πιο εύκολο να ερμηνευτούν και να κατανοηθούν από τους ενδιαφερόμενους, διευκολύνοντας τις πρωτοβουλίες βελτίωσης της διαδικασίας. Σημαντικό, εξαιτίας των χαρακτηριστικών που έχουν τα δέντρα διαδικασίας, είναι ότι υποστηρίζει την ανακάλυψη παράλληλης συμπεριφοράς και γενικά την πολυπλοκότητα των διαδικασιών. Τέλος, η επαναληπτική φύση του αλγορίθμου κάνει την εργασία βελτίωσης και την προσαρμογής του μοντέλου διεργασίας ευκολότερη, ενισχύοντας την ακρίβεια και την πληρότητα.

### 3.4.3 Heuristic Miner

Ο Ευρετικός αλγόριθμος (Heuristic Miner) αποτελεί μια βελτιωμένη έκδοση του αλγορίθμου- $a$  η οποία λαμβάνει σαν κύριο χαρακτηριστικό την συχνότητα, λαμβάνοντας σαν είσοδο πέραν των σχέσεων μεταξύ των δραστηριοτήτων και το πόσες φορές μία δραστηριότητα ακολουθείται από μία άλλη μέσα στο αρχείο καταγραφής γεγονότων. Στην συνέχεια, για κάθε σχέση υπολογίζεται η μεταξύ τους εξάρτηση και δημιουργείται ένας πίνακας εξαρτήσεων που καταγράφει όλες τις σχέσεις των δραστηριοτήτων της διαδικασίας και αποτελεί την είσοδο του Ευρετικού Αλγορίθμου. Αυτός ο πίνακας προσδιορίζει τις σχέσεις προτεραιότητας μεταξύ δραστηριοτήτων με βάση τη σειρά με την οποία εμφανίζονται στο αρχείο καταγραφής συμβάντων. Αναλύοντας τις εξαρτήσεις, το Heuristic Miner μπορεί να καθορίσει την ακολουθία των δραστηριοτήτων που συνήθως συμβαίνουν μαζί. Παράλληλα καθορίζεται και ένα κατώφλι (threshold) με το οποίο οι δραστηριότητες που έχουν καταγραφεί και είναι κάτω από αυτό δεν εμφανίζονται στο μοντέλο διαδικασίας. Με αυτόν τον τρόπο μπορεί να γίνει απαλοιφή των σπάνιων περιπτώσεων που βρίσκονται στο αρχείο καταγραφής και δημιουργούν ένα είδος θορύβου που δεν μπορεί να διαχειριστεί ο αλγόριθμος- $a$ .

Παρακάτω ο τύπος που υπολογίζεται, η τιμή της σχέσης εξάρτησης μεταξύ των δραστηριοτήτων  $a$  και  $b$ :

$$|a \Rightarrow_L b| = \frac{|a >_L b| - |b >_L a|}{|a >_L b| + |b >_L a| + 1} \quad \text{if } a \neq b$$

$$|a \Rightarrow_L b| = \frac{|a >_L a|}{|a >_L a| + 1} \quad \text{if } a = b$$

Ο Heuristic Miner πλεονεκτεί έναντι των υπολοίπων ως προς την επέκτασή και την χρήση μεγάλων συνόλων δεδομένων και να επεξεργάζεται αποτελεσματικά τα δεδομένα του αρχείου καταγραφής συμβάντων και παράγει αξιόπιστα αποτελέσματα ακόμη και όταν αντιμετωπίζει πολύπλοκα μοντέλα διεργασιών. Επιπλέον, μπορεί να χειριστεί διαφορετικές παραλλαγές της διαδικασίας και δεν περιορίζεται από άκαμπτα μοντέλα. Αυτό το καθιστά κατάλληλο για παραγωγικές διαδικασίες όπου ενδέχεται να παρουσιάζουν δυναμική συμπεριφορά. Τέλος, αυτοματοποιεί τις εργασίες ανακάλυψης και ελέγχου διαδικασιών, μειώνοντας την ανάγκη για χειροκίνητη παρέμβαση. Επιταχύνει τη διαδικασία ανάλυσης και ελαχιστοποιεί το ανθρώπινο λάθος, καθιστώντας το πολύτιμο εργαλείο για οργανισμούς που στοχεύουν στη βελτιστοποίηση των διαδικασιών τους.

### 3.5 Εργασίες εξόρυξης διαδικασιών

Έχει αναφερθεί ότι η εξόρυξη διαδικασιών είναι μέρος των σταδίων της ανάλυσης και του επανασχεδιασμού στον κύκλο ζωής ενός έργου διοίκησης επιχειρηματικών διαδικασιών (BPM-life cycle). Ένα αντίστοιχο έργο εξόρυξης διαδικασιών, επίσης, αποτελείται από διαφορετικά στάδια τα οποία εκτελούνται και είναι αποτέλεσμα των αντίστοιχων τεχνικών. Στην βιβλιογραφία [19] αναφέρονται σαν τύποι εξόρυξης διαδικασιών (Process Mining Types). Αλλά πριν γίνει αναφορά στους τύπους πρέπει να γίνουν κατανοητές οι τεχνικές που εφαρμόζονται μεταξύ των μοντέλων διαδικασιών και των αρχείων καταγραφής δεδομένων και η σύνδεσή τους με τους αντίστοιχους τύπους εξόρυξης διαδικασιών.

Οι σχέσεις και οι αντίστοιχες τεχνικές μεταξύ αρχείων καταγραφής συμβάντων που έχουν θεσπιστεί στο πλαίσιο της εξόρυξης δεδομένων βοηθώντας στην ανάλυση και οπτικοποίηση της συμπεριφοράς των επιχειρηματικών διαδικασιών για τον εντοπισμό περιοχών προς βελτίωση είναι οι Play-In, Play-Out και Replay. Αρχικά, το *Play-in* είναι μια τεχνική εξόρυξης διεργασιών που σαν είσοδο έχει τα αρχεία καταγραφής συμβάντων και σαν έξοδο το μοντέλο διαδικασιών με το οποίο γίνεται ανάλυση της συμπεριφοράς μιας επιχειρηματικής διαδικασίας καθώς αυτή εκτυλίσσεται σε πραγματικό χρόνο. Αυτή η τεχνική μπορεί να χρησιμοποιηθεί για τον εντοπισμό σημείων συμφόρησης, αναποτελεσματικότητας και άλλων ζητημάτων που μπορεί να επηρεάζουν την απόδοση της διαδικασίας. Αναλύοντας με αυτόν το τρόπο τα αρχεία καταγραφής συμβάντων, οι οργανισμοί μπορούν να αναλάβουν δράση για την αντιμετώπιση αυτών των ζητημάτων όταν προκύπτουν. Το *Play-out* είναι μια τεχνική εξόρυξης διεργασιών που έχει σαν είσοδο ένα μοντέλο διαδικασιών σε μορφή PetriNet και σαν έξοδο ένα αρχείο καταγραφής συμβάντων. Αυτή η τεχνική μπορεί να χρησιμοποιηθεί για τον εντοπισμό προτύπων, τάσεων και άλλων γνώσεων σχετικά με τη συμπεριφορά της διαδικασίας. Μπορεί να χρησιμοποιηθεί για εργασίες προσομοίωσης συλλέγοντας και αναλύοντας στοιχεία από τα αρχεία καταγραφής συμβάντων που θα προκύψουν μετά την ολοκλήρωση της διαδικασίας ώστε να κατανοηθεί καλύτερα ο τρόπος εκτέλεσης της και να εντοπιστούν τα μέρη εκείνα που χρειάζονται βελτίωση. Τέλος, το *Replay* είναι μια τεχνική που σαν είσοδο χρησιμοποιεί ταυτόχρονα το μοντέλο διαδικασιών που έχει δημιουργηθεί και τα αρχεία καταγραφής συμβάντων με στόχο την προσομοίωση της διαδικασίας. Σαν αποτέλεσμα έχει εκτενή μοντέλα διαδικασιών με περισσότερες πληροφορίες. Επίσης μπορεί να πραγματοποιηθεί για έλεγχο συμμόρφωσης, προτάσεις για βελτίωση της διαδικασίας, να γίνει εξαγωγή μοντέλων που θα προβλέπουν συμπεριφορές της και συγκεκριμένους διαγνωστικούς ελέγχους στο πλαίσιο της εξόρυξης

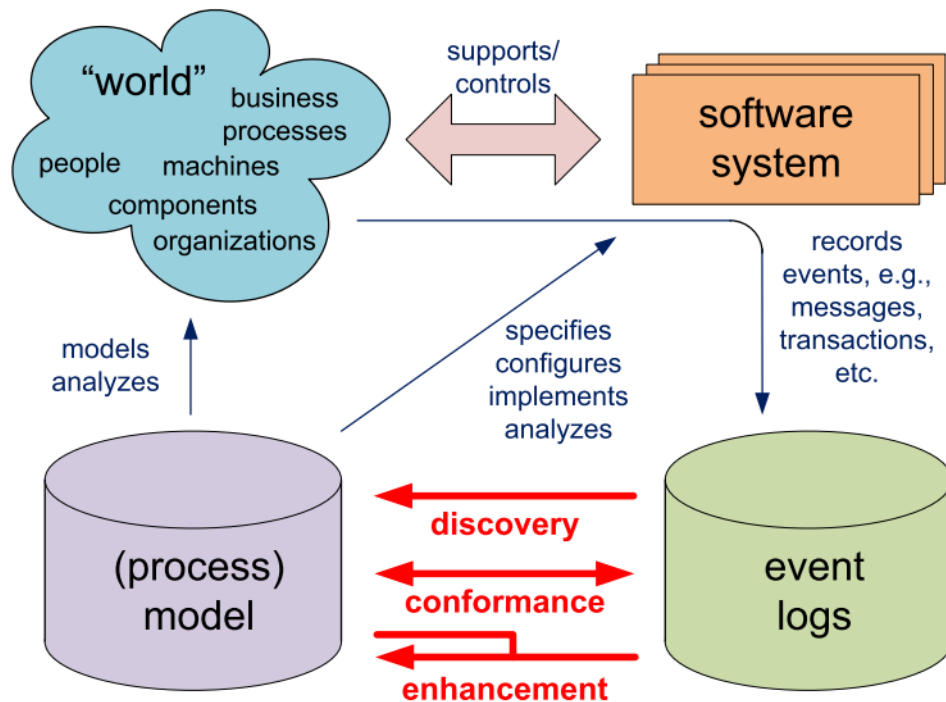


Figure 3.10: Κύριοι τύποι εξόρυξης διεργασιών: discovery, conformance και enhancement. Πηγή [21]

διαδικασιών.

Έπειτα από την ανάλυση και επεξήγηση σχετικά με τα αρχεία καταγραφής συμβάντων, τα μοντέλα διαδικασιών και τις τεχνικές με τις οποίες σχετίζονται, παρουσιάζονται παρακάτω κάποιες εργασίες ενός έργου εξόρυξης διαδικασιών. Βασικοί τύποι είναι η ανακάλυψη της διαδικασίας (Process Discovery), ο έλεγχος συμμόρφωσης (Conformance Checking) και η βελτιστοποίηση της διαδικασίας (Enhancement). Επιπρόσθετες εργασίες είναι η ανάλυση απόδοσης (Performance Analysis), η συγκριτική εξόρυξη διαδικασιών (Comparative Process Mining), η Προγνωστική Εξόρυξη Διαδικασιών (Predictive Process Mining) και η εξόρυξη διαδικασιών με προσανατολισμό τα actions (Action-Oriented Process Mining).

### 3.5.1 Ανακάλυψη της διαδικασίας (Process Discovery).

Η ανακάλυψη της διαδικασίας είναι η αρχική και μια από τις πιο απαιτητικές εργασίες στο πλαίσιο ενός έργου εξόρυξης διαδικασιών. Από τα αρχεία καταγραφής συμβάντων γίνεται η ανακάλυψη του μοντέλου διαδικασιών, σε συνδυασμό με την χρήση κάποιου αλγόριθμου εξόρυξης διαδικασιών όπως είναι οι A-miner, inductive miner, heuristic miner κα, με σκοπό να εξαχθεί το ιδανικό μοντέλο. Σύμφωνα με την κατηγοριοποίηση των τεχνικών εξόρυξης διαδικασιών πρόκειται για μία *Play-In* τεχνική που έχει σαν είσοδο ένα αρχείο καταγραφής συμβάντων και σαν έξοδο ένα μοντέλο διαδικασίας (BPMN, PetriNet, CasualNet, κ.α.). Η εργασία πέραν της επιλογής του κατάλληλου αλγορίθμου, περιλαμβάνει την ανακάλυψη και τον χαρακτηρισμό των δραστηριοτήτων, την ανακάλυψη της ροής της διαδικασίας και τους πόρους που χρησιμοποιεί. Μια απλή προσέγγιση είναι η δημιουργία ενός γράφου, Directly Follow Graph (DFG) στον οποίο

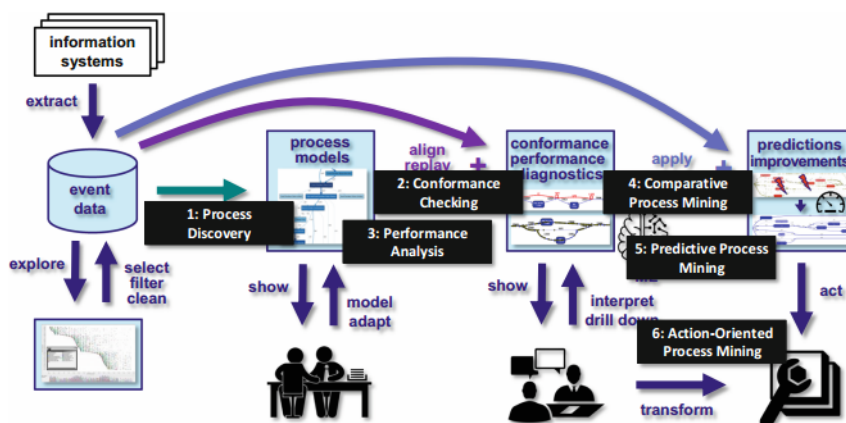


Figure 3.11: Συχνές εργασίες ενός έργου εξόρυξης διαδικασιών. Πηγή

οι δραστηριότητες είναι οι κόμβοι οι οποίοι συνδέονται μεταξύ τους αν οι δραστηριότητες ακολουθούν η μία την άλλη [13]. Πρόκειται, όμως, στο σύνολο της για αρκετά δύσκολη εργασία με πολλές προκλήσεις. Για παράδειγμα, πρέπει να αποφασιστεί αν πρέπει να συμπεριληφθούν συμπεριφορές που είναι αρκετά σπάνιες και καταγράφονται στα event logs, και αν είναι κάτι το οποίο θα περιπλέξει το μοντέλο ή είναι μια συμπεριφορά η οποία είναι σημαντικό να καταγραφεί.

Με τον όρο "ανακάλυψη της διαδικασίας" κυρίως αναφερόμαστε στην ανακάλυψη της ροής εκτέλεσης των διαδικασιών (control-flow perspective) η οποία επικεντρώνεται στην εξαγωγή ενός μοντέλου το οποίο παρουσιάζει την χρονολογική εκτέλεση των δραστηριοτήτων, εφόσον τα γεγονότα στο αρχείο καταγραφής το οποίο έχουμε σαν είσοδο αντιστοιχούν σε εκτελέσεις δραστηριοτήτων στο σύνολο της διαδικασίας που αναλύεται. Γίνεται προσπάθεια να κατασκευαστεί ένα μοντέλο που να αντιπροσωπεύει την πιο συχνή ροή που παρατηρείται στο αρχείο καταγραφής. Οι βασικές τεχνικές για την συγκεκριμένη εργασία είναι, ουσιαστικά, οι αλγόριθμοι ανα διαδικασιών σε συνδυασμό με τεχνικές που εντάσσονται στην επιστήμη της εξόρυξης δεδομένων. Η επιλογή της τεχνικής ανακάλυψης διαδικασίας εξαρτάται από συγκεκριμένα χαρακτηριστικά του αρχείου καταγραφής, τις απαιτήσεις της επιχειρησιακής διαδικασίας και την αναπαράσταση του μοντέλου που παράγεται.

### 3.5.2 Έλεγχος συμμόρφωσης (Conformance Checking).

Ο έλεγχος συμμόρφωσης (Conformance checking) είναι μια από τις κύριες εργασίες σε ένα έργο εξόρυξης διαδικασιών και, σύμφωνα με την κατηγοριοποίηση των τεχνικών, πρόκειται για μια *Replay* τεχνική που σαν είσοδο έχει ένα μοντέλο διαδικασίας (PetriNet, DFG, BPMN κλπ) και ένα αρχείο καταγραφής γεγονότων. Βασική ιδέα είναι η σύγκριση της μοντελοποιούμενης συμπεριφοράς που αναπαριστά το μοντέλο διαδικασίας με την παρατηρούμενη συμπεριφορά η οποία καταγράφεται στα αρχεία καταγραφής δεδομένων. Συσχετίζοντας αυτές τις δύο οντότητες στόχος είναι να αναλυθεί και να κατανοηθεί το μεταξύ τους επίπεδο συμμόρφωσης ανακαλύπτοντας με διάφορες τεχνικές τα κοινά τους σημεία, τις διαφορές και τις αποκλίσεις τους.

Τα αποτελέσματα του ελέγχου συμμόρφωσης ποικίλουν και οι χρήσεις τους είναι πολλαπλές με την φύση της διαδικασίας και του εννοιολογικού πλαισίου στο οποίο εκτελείται. Σε τεχνικό επίπεδο, σε ότι αφορά το πλαίσιο της ανάλυσης, το αποτέλεσμα



είναι αρχικά, κατά πόσο το ανακαλυφθέν μοντέλο διαδικασίας μπορεί να αναπαράγει ορθά τα ίχνη από τα οποία δημιουργήθηκε ή ίχνη της ίδιας διαδικασίας που δεν χρησιμοποιήθηκαν στην κατασκευή του και η μετέπειτα επιδιόρθωση του αν αυτό δεν ευθυγραμμίζεται με την πραγματικότητα. Επιπλέον, χρησιμοποιείται στο πλαίσιο μέτρησης της απόδοσης των αλγορίθμων εξόρυξης διαδικασιών και το κατά πόσο το μοντέλο που παράγουν, είναι κοντά στην παρατηρούμενη συμπεριφορά.

Σε παραγωγικό επίπεδο ο έλεγχος συμμόρφωσης αποτελεί ένα δυνατό εργαλείο για τους οργανισμούς. Είναι σημαντικός για την συμμόρφωση, ευθυγράμμιση και των ελέγχου τον επιχειρηματικών διαδικασιών. Ελέγχει την διαδικασία από τα αρχικά στάδια της δημιουργίας της και εξασφαλίζει χρόνο επιλύοντας άμεσα διαφορές και αποκλίσεις. Κατά την ευθυγράμμιση των επιχειρηματικών διαδικασιών (business alignment<sup>8</sup>) το conformance checking διασφαλίζει την αρμονική συνεργασία των πληροφοριακών συστημάτων(πχ SAP, Jira) και των χειριστών τους με τις επιχειρηματικές διαδικασίες. Τέλος, κατά τον έλεγχο(auditing) γίνεται η αξιολόγηση του οργανισμού και των διαδικασιών τους με γνώμονα την εσωτερική πολιτική της εταιρίας ή τους κρατικούς και κοινοτικούς κανόνες και οδηγίες.

Υπάρχουν διάφορες τεχνικές ελέγχου συμμόρφωσης που χρησιμοποιούνται στην εξόρυξη διαδικασιών οι οποίες αξιοποιώντας τα δεδομένα των αρχείων καταγραφής και τα μοντέλα διαδικασιών ενισχύουν τις πρωτοβουλίες βελτίωσης διαδικασιών. Παρακάτω παρουσιάζονται περιγραφικά οι κυρίαρχες τεχνικές:

- *Token Replay*. Η τεχνική token-based replay συγκρίνει ένα ίχνος που βρίσκεται σε ένα αρχείο καταγραφής γεγονότων και ένα μοντέλο PetriNet και βασικό στόχο της αποτελεί η ποσοτικοποίηση καταλληλότητάς(fitness). Μια απλή προσέγγιση για τον υπολογισμό της είναι να μετρήσει ποια ίχνη του αρχείου καταγραφής μπορούν να αναπαραχθούν από το μοντέλο και ποια όχι και να βρεθούν τα αντίστοιχα ποσοστά καταλληλότητας. Η τεχνική token replay όμως, εκμεταλλεύεται το γεγονός παραγωγής token από ένα PetriNet και υπολογίζει το fitness σε επίπεδο events. Ξεκινώντας από την αρχική θέση, κάθε μετάβαση πυροδοτείται ανεξαρτήτως αν υπάρχει token ή όχι, προκειμένου στο τέλος να γίνει αξιολόγησή και να ανακαλύψει ποιες μεταβάσεις εκτελέστηκαν και σε ποιες θέσεις έχουμε εναπομείναντα(remaining) tokens, που απουσιάζουν(missing) και πόσα από αυτά έφτασαν στο τέλος για μια συγκεκριμένη διαδικασία. Για το τελικό αποτέλεσμα και την αξιολόγηση ο αναλυτής μπορεί να δημιουργήσει μια κατηγοριοποίηση των traces με αυτά , που ταιριάζουν 100%(fitting), με αυτά που σχεδόν ταιριάζουν(almost fitting) και εκείνα που δεν ταιριάζουν καθόλου(not fitting).

Τύπος υπολογισμού καταλληλότητας με Token-replay για κάθε ίχνος:

$$\text{Fitness}(\sigma, N) = \frac{1}{2} \left(1 - \frac{m}{c}\right) + \frac{1}{2} \left(1 - \frac{r}{p}\right)$$

<sup>8</sup>Η επιχειρησιακή ευθυγράμμιση (business alignment) αναφέρεται στην επίτευξη συμφωνίας μεταξύ των στρατηγικών στόχων και των δραστηριοτήτων μιας επιχείρησης. Πρόκειται για τον συντονισμό και την αρμονία μεταξύ της στρατηγικής κατεύθυνσης της επιχείρησης και των δράσεων που αναλαμβάνονται για την επίτευξη αυτών των στόχων. Μια απόδειξη επιχειρησιακής ευθυγράμμισης είναι όταν οι διαδικασίες, οι δραστηριότητες και οι αποφάσεις που λαμβάνονται στην επιχείρηση είναι αρμονικά ευθυγραμμισμένες με τους στρατηγικούς στόχους και τις αξίες της. Η επιχειρησιακή ευθυγράμμιση επιτρέπει στην επιχείρηση να επιτύχει αποτελεσματικότητα, ανταγωνιστικότητα και βιωσιμότητα.[12],[14]

Τύπος υπολογισμού με Token-replay Fitness συνολικά για το Event Log:

$$\text{Fitness}(L, N) = \frac{1}{2} \left( 1 - \frac{\sum_{\sigma \in L} L(\sigma) \times mN, \sigma}{\sum_{\sigma \in L} L(\sigma) \times cN, \sigma} \right) + \frac{1}{2} \left( 1 - \frac{\sum_{\sigma \in L} L(\sigma) \times rN, \sigma}{\sum_{\sigma \in L} L(\sigma) \times pN, \sigma} \right)$$

Μειονεκτήματα της συγκεκριμένης τεχνικής είναι ότι αν υπάρχουν πολλές αποκλίσεις. Τα PetriNets πλημμυρίζουν με tokens, με αποτέλεσμα να επιτρέπουν οποιαδήποτε συμπεριφορά. Η προσέγγιση είναι αποκλειστικά εξαρτημένη από το μοντέλο και αν ένα case δεν ταιριάζει, η τεχνική αυτή δεν δημιουργεί μία αντίστοιχη εναλλακτική διαδρομή, και ο έλεγχος αποτυγχάνει, ουσιαστικά προκαταβάλλει την διαδρομή και δεν είναι δυνατή η μέτρηση στο σύνολο του μοντέλου.

- *Alignments.* Η τεχνική βασισμένη στην ευθυγραμμίσεις έχει σαν στόχο στον εντοπισμό της καλύτερης ευθυγράμμισης μεταξύ του ίχνους και ενός μοντέλου διαδικασιών (PetriNet, Process Tree, DFG κλπ). Για κάθε ίχνος, το αποτέλεσμα μιας ευθυγράμμισης είναι μία λίστα από ζεύγη, όπου το πρώτο στοιχείο είναι ένα event από το trace ή το σύμβολο "»" και το δεύτερο στοιχείο είναι μια μετάβαση του μοντέλου ή το σύμβολο "»". Με την συγκεκριμένη μέθοδο παρέχονται λεπτομερή διαγνωστικά για κάθε περίπτωση, τα οποία μπορούν να συγκεντρωθούν και να υπολογιστούν συνολικά στο μοντέλο. Είναι, λοιπόν, εύκολο να εντοπιστεί αν κάποια δραστηριότητα συχνά παραλείπεται ή ότι κάποια συμβαίνει ενώ δεν θα έπρεπε. Σημαντικό είναι να αναφερθεί ότι η καλύτερη ευθυγράμμιση επιλέγεται βάσει καλύτερου κόστους το οποίο αναφέρεται στο μέτρο της απόκλισης μεταξύ της παρατηρούμενης συμπεριφοράς που καταγράφεται σε ένα αρχείο και της αναμενόμενης συμπεριφοράς που μοντελοποιείται σε ένα μοντέλο διαδικασίας. Αποτελεί τη μέτρηση του βαθμού στον οποίο η παρατηρούμενη συμπεριφορά αποκλίνει από την ιδανική ή αναμενόμενη συμπεριφορά που ορίζεται από το μοντέλο διαδικασίας. Επίσης, βάσει του κόστους υπολογίζεται και η καταλληλότητα (fitness) τόσο σε επίπεδο ίχνους όσο και σε επίπεδο αρχείου καταγραφής με τους παρακάτω τύπους.

Τύπος υπολογισμού καταλληλότητάς (Fitness) με ευθυγραμμίσεις ξεχωριστά για κάθε ίχνος:

$$\text{fitness}(\sigma, N) = 1 - \frac{\delta(\lambda_{\text{opt}}(N, \sigma))}{\delta(\lambda_{\text{worst}}(N, \sigma))}$$

Τύπος υπολογισμού καταλληλότητάς (Fitness) με ευθυγραμμίσεις συνολικά για το Event Log:

$$\text{fitness}(L, N) = 1 - \frac{\sum_{\sigma \in L} L(\sigma) \times \delta(\lambda_{\text{opt}}(N, \sigma))}{\sum_{\sigma \in L} L(\sigma) \times \delta(\lambda_{\text{worst}}(N, \sigma))}$$

	A_ACCEPTED	A_ACTIVATED	A_APPROVED	A_CANCELLED	A_DECLINED	A_FINALIZED	A_PARTLYSUBMITTED	A_PREACCEPTED
A_ACCEPTED	#	#	#	#	#	>	#	<
A_ACTIVATED	#	#		#	#	<	#	#
A_APPROVED	#		#	#	#	<	#	#
A_CANCELLED	#	#	#	#	#	<	#	<
A_DECLINED	#	#	#	#	#	<	#	#
A_FINALIZED	<	>	>	>	>	#	#	#
A_PARTLYSUBMITTED	#	#	#	#	#	#	#	>
A_PREACCEPTED	>	#	#	>	#	#	<	#
A_REGISTERED	#			#	#	<	#	#
A_SUBMITTED	#	#	#	#	#	#	>	#

Figure 3.12: Παράδειγμα ενός FootPrint το οποίο παράγει η Pm4py.

- *Footprints*. Ο έλεγχος συμμόρφωσης βασισμένος στα Footprints είναι μια βασική και κλιμακούμενη τεχνική η οποία συγκρίνει τις οντότητες (αρχεία καταγραφής, PetriNet, DFG, Process Tree κλπ) που συμμετέχουν στην εξόρυξη διαδικασιών μεταξύ τους. Ουσιαστικά πρόκειται για δισδιάστατους πίνακες οι οποίοι αποτυπώνουν την σχέση ενός ζευγαριού δραστηριοτήτων που συμμετέχουν στην διαδικασία. Βασική ιδέα είναι να δημιουργηθούν όλοι οι πίνακες δυο διαστάσεων και να συγκριθούν μεταξύ τους έτσι ώστε να βρεθούν τα σημεία που συμπίπτουν και τα σημεία στα οποία υπάρχουν αποκλίσεις, ώστε να υπολογιστεί η μεταξύ τους καταλληλότητα. Οι μορφές των σχέσεων μεταξύ των δραστηριοτήτων από την βιβλιογραφία έχει καθοριστεί ως εξής,
  - Directly-Follows Relationships, όταν η δραστηριότητα A ακολουθείται απευθείας από την δραστηριότητα B. (A->B)
  - Directly-Before Relationships, όταν η δραστηριότητα B προηγείται της δραστηριότητα A. (B->A)
  - Parallel behavior, όταν η δραστηριότητα B ακολουθείται από την δραστηριότητα A και η δραστηριότητα A ακολουθείται από την δραστηριότητα B. (B||A)

Κάθε μία από τις παραπάνω τεχνικές καλύπτει και υποστηρίζει την ανακάλυψη λαθών και τον εντοπισμό τους αν αυτά υπάρχουν στο μοντέλο ή στην διαδικασία. Ποσοτικοποιούν την συμμόρφωση και κάνουν διάγνωση της μη συμμόρφωσης. Επίσης, πρέπει να γίνει κατανοητό ότι ο έλεγχος συμμόρφωσης και η ιδιότητά του να ποσοτικοποιήσει την ικανότητα του μοντέλου να αναπαράγει την παρατηρούμενη συμπεριφορά μπορεί να χρησιμοποιηθεί για αξιολόγηση των αλγορίθμων εξόρυξης διαδικασιών.

### 3.5.3 Βελτιστοποίηση (Enhancement).

Με τις εργασίες της ανακάλυψης της διαδικασίας και του ελέγχου συμμόρφωσης η διαδικασία, κυρίως, ελέγχεται από την οπτική της ροής εκτέλεσης των δραστηριοτήτων. Η βελτιστοποίηση αφορά τον εμπλουτισμό του μοντέλου διαδικασίας με δεδομένα της διαδικασίας που είναι διαθέσιμα με σκοπό να βελτιώσει ή να επιδιορθώσει το αρχικό μοντέλο και να ταιριάζει περισσότερο στην πραγματικότητα. Η εργασία αυτή εκμεταλλεύεται την ιδιότητα των αρχείων καταγραφής γεγονότων να μπορούν να αποθηκεύσουν περισσότερες πληροφορίες για την διαδικασία και την ροή της, όπως για παράδειγμα το ποιος εκτέλεσε μια συγκεκριμένη δραστηριότητα ή τις ιδιότητές του πελάτη που εξυπηρετείται.

Μέρος της βελτιστοποίησης είναι η χρήση τεχνικών που έχουν ενσωματωθεί από άλλους τομείς όπως το data mining. Για παράδειγμα, εφαρμόζεται ανάλυση mining decision point για να ανακαλυφθεί τί ήταν αυτό που οδήγησε σε μια συγκεκριμένη

απόφαση που άλλαξε την ροή της εκτέλεσης και των διαδικασιών. Αναλύει την συμπεριφορά των πόρων, γι' αυτό εκτελεί τεχνικές που σχετίζονται με την ανάλυση κοινωνικών δικτύων. Τέλος, μπορεί να εξάγει την απόδοση των διαδικασιών, οπότε συσχετίζει τις τεχνικές από την ερευνητική λειτουργική ανάλυση, όπως για παράδειγμα τον προσομοιωτή ή τα δίκτυα ουρών για τεχνικές εξόρυξης διαδικασιών.

### 3.5.4 Ανάλυση απόδοσης (Performance Analysis).

Μέσα από την εργασία της Ανάλυσης Απόδοσης γίνεται επίτευξη του στόχου των εργασιών εξόρυξης διαδικασιών που είναι να βελτιώσει τις διαδικασίες φανερώνοντας τα προβλήματα. Αυτά μπορεί να είναι τα προβλήματα συμμόρφωσης που περιγράφηκαν, προβλήματα απόδοσης, όπως η μη έγκαιρη ολοκλήρωση μιας περίπτωσης, την περιορισμένη εκτέλεση δραστηριοτήτων, καθυστερήσεις, επαναλήψεις της ίδιας διαδικασίας και προβλήματα ποιότητας. Χρησιμοποιώντας την αναπαραγωγή τις τεχνικές footprints και alignments είναι δυνατόν να αντιστοιχηθούν τα δεδομένα του αρχείου καταγραφής γεγονότων με ένα μοντέλο διαδικασίας. Ως αποτέλεσμα, είναι αρκετά απλό να εμπλουτιστεί το μοντέλο διαδικασίας με πληροφορίες συχνότητας και χρόνου. Οι συχνότητες των μη επιθυμητών δραστηριοτήτων και βρόχων μπορούν να χρησιμοποιηθούν για την αναγνώριση προβλημάτων καθυστέρησης και αποδοτικότητας. Αφού τα γεγονότα έχουν χρονικά στιγμιότυπα, είναι δυνατόν να μετρηθούν οι χρόνοι μεταξύ των δραστηριοτήτων, συμπεριλαμβανομένων στατιστικών όπως ο μέσος όρος, η τυπική απόκλιση, η ελάχιστη και μέγιστη χρονική απόδοση μιας περίπτωσης διαδικασίας. Αυτό επιτρέπει την ανάλυση των δεικτών απόδοσης, όπως οι χρόνοι αναμονής, οι χρόνοι απόκρισης και οι χρόνοι εξυπηρέτησης.

Παράδειγμά ανάλυσης απόδοσης αποτελεί η ανάλυση των SLA σε ένα λογισμικό διαχείρισης έργων και πελατών όπως είναι το Jira. Τα SLA (Συμφωνηθέν Επίπεδο Υπηρεσίας) είναι μια συμφωνία μεταξύ ενός παρόχου υπηρεσιών και ενός πελάτη. Το process mining μπορεί να χρησιμοποιηθεί για την ανάλυση των SLA, και να ανακαλύψει πότε δεν εκπληρώνεται ένα συγκεκριμένο SLA. Ορισμένα γνωστά SLA είναι ο αριθμός χαμένων περιπτώσεων, ο μέσος χρόνος απάντησης (χρόνος απόκρισης που βλέπει ο πελάτης), το ποσοστό περιπτώσεων που επεξεργάστηκαν εντός ενός προκαθορισμένου χρονικού πλαισίου, η επίλυση με την πρώτη κλήση (περιπτώσεις που επεξεργάστηκαν με επιτυχία χωρίς ανακατεύθυνση), το ποσοστό διπλοτύπων περιπτώσεων (π.χ., πολλά έγγραφα προμήθειας που αντιστοιχούν στην ίδια παραγγελία), ο μέσος χρόνος μεταξύ αποτυχιών, ο μέσος χρόνος ανάκαμψης κλπ.[22]



# Εργαλεία Εξόρυξης Διαδικασιών

Με τον όρο "λογισμικό ανάλυσης και εξόρυξης διαδικασιών" η βιβλιογραφία αναφέρεται στα εργαλεία και τις εφαρμογές που χρησιμοποιούνται για την εκτέλεση εργασιών ανάλυσης και εξόρυξης διαδικασιών. Τα βασικά χαρακτηριστικά που πρέπει να έχει μια εφαρμογή που εκτελεί εργασίες εξόρυξης διαδικασιών, είναι η εισαγωγή και προεπεξεργασία των αρχείων καταγραφής γεγονότων, η υποστήριξη ενός συνόλου αλγορίθμων ανακάλυψης διαδικασιών, η υποστήριξη κατασκευής ενός πλήθους μοντέλων διαδικασιών, η δυνατότητα ελέγχου συμμόρφωσης και λειτουργιών βελτίωσης διαδικασιών. Προηγμένα εργαλεία μπορεί επίσης να παρέχουν δυνατότητες προβλεπτικής ανάλυσης, παρακολούθησης της διαδικασίας σε πραγματικό χρόνο, σύνδεση με άλλα εταιρικά εργαλεία όπως είναι οι διακομιστές ηλεκτρονικών μηνυμάτων (Outlook), συστήματα επιχειρηματικού σχεδιασμού - ERP (πχ SAP), προγράμματα περιήγησης (Firefox, Chrome), repositories (Github) κ.α.

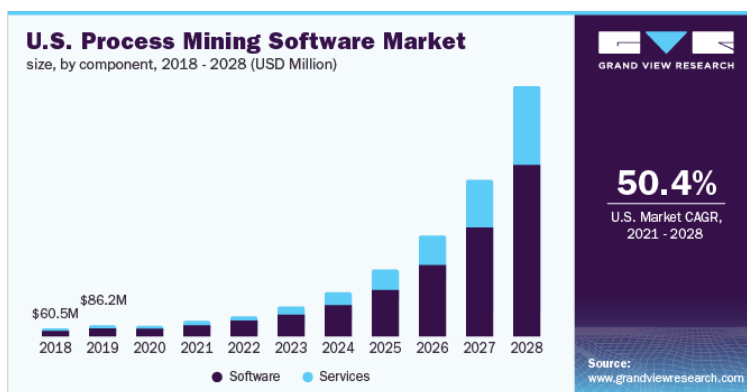


Figure 4.1: Μερίδιο της Αμερικάνικης αγοράς λογισμικού εξόρυξης διαδικασιών. Πηγή[11]

Αξίζει να αναφερθεί ότι ο ψηφιακός μετασχηματισμός, η ανάγκη των οργανισμών για έλεγχο, κατανόηση και ανάλυση των επιχειρηματικών διαδικασιών καθώς και η συμμόρφωσή αυτών είτε με επιχειρησιακούς είτε με κρατικούς κανονισμούς ωθεί εταιρείες στην επένδυση σε λογισμικό εξόρυξης διαδικασιών είτε ανεξάρτητο, είτε στο πλαίσιο ενός εργαλείου διαχείρισης επιχειρησιακών διαδικασιών, είτε στην υιοθέτηση λύσεων νέφους (cloud). Σύμφωνα με έρευνα[11] η παγκόσμια αγορά λογισμικού εξόρυξης διαδικασιών υπολογίζεται στα 322,02 εκατομμύρια δολάρια το 2020 και αναμένεται ραγδαία αύξησή της με ρυθμό 50.1% από το 2021 έως το 2028. Στις παρακάτω ενότητες παρουσιάζονται δύο από τα πιο διαδεδομένα εργαλεία εξόρυξης διαδικασιών, το ProM και το Disco καθώς και η βιβλιοθήκη της γλώσσας προγραμματισμού της Python, *pm4py*.

Table 4.1: Εργαλεία εξόρυξης διεργασιών που είναι διαθέσιμα στο τέλος του 2021. Για κάθε εργαλείο παρατίθεται ο προμηθευτής και ο ιστότοπος του. Η τελευταία στήλη υποδεικνύει εάν υπάρχει διαθέσιμη ακαδημαϊκή έκδοση[24].

Ventor	Tool	Website	Acad. ver
Abbyy	ABBYY Timeline	<a href="http://www.abbyy.com">www.abbyy.com</a>	No
Appian (Lana Labs)	LANA Process Mining	<a href="http://appian.com">appian.com</a>	No
Apomore	Apomore Enterprise Edition	<a href="http://apomore.org">apomore.org</a>	Yes
bupaR	bupaR	<a href="http://bupar.net">bupar.net</a>	Yes
businessOptix	businessOptix	<a href="http://businessoptix.com">businessoptix.com</a>	Yes
Celonis	Celonis EMS	<a href="http://celonis.com">celonis.com</a>	Yes
Datricks	Datricks	<a href="http://datricks.com">datricks.com</a>	Yes
DCR	DCR Portal	<a href="http://dcrsolutions.net">dcrsolutions.net</a>	Yes
Deloitte	Process X-ray	<a href="http://processxray.deloitte.com">processxray.deloitte.com</a>	No
EverFlow	EverFlow	<a href="http://everflow.io">everflow.io</a>	No
Fluxicon	Disco	<a href="http://fluxicon.com">fluxicon.com</a>	Yes
FortressIQ	FortressIQ	<a href="http://fortressiq.com">fortressiq.com</a>	No
Fraunhofer FIT	PM4py	<a href="http://pm4py.fit.fraunhofer.de">pm4py.fit.fraunhofer.de</a>	Yes
Hyland	Onbase	<a href="http://hyland.com">hyland.com</a>	No
IBM(myInvenio)	myInvenio	<a href="http://my-invenio.com">my-invenio.com</a>	No
Integris	Explora	<a href="http://integris.it">integris.it</a>	No
Kofax	Kofax Insight	<a href="http://www.kofax.com">www.kofax.com</a>	No
livejourney	livejourney	<a href="http://www.livejourney.com">www.livejourney.com</a>	No
Logpickr	Process Explorer 360	<a href="http://www.logpickr.com">www.logpickr.com</a>	No
Mavim	Mavim	<a href="http://www.mavim.com">www.mavim.com</a>	No
Mehrwerk GmbH	MPM	<a href="http://mpm-processmining.com">mpm-processmining.com</a>	No
Mindzie	mindzie	<a href="http://mindzie.com">mindzie.com</a>	Yes
Minit(Microsoft)	Minit	<a href="http://www.minit.io">www.minit.io</a>	Yes
Nintex UK ltd	Nintex	<a href="http://www.nintex.com">www.nintex.com</a>	No
Oniq	IQ/A	<a href="http://www.oniq.com">www.oniq.com</a>	No
PAFnow (Celonis)	PAFnow	<a href="http://www.pafnow.com">www.pafnow.com</a>	No
Process.science	process.science	<a href="http://www.process-science.com">www.process-science.com</a>	No
ProcessDiamond	ProcessDiamond	<a href="http://www.decisions.com">www.decisions.com</a>	Yes
ProcessM	PmBI	<a href="http://www.processm.com">www.processm.com</a>	Yes
Puzzle Data	ProDiscovery	<a href="http://www.puzzledata.com">www.puzzledata.com</a>	No
QPR Software	QPR ProcessAnalyzer	<a href="http://www.qpr.com">www.qpr.com</a>	No
SAP (Signavio)	SAP Signavio	<a href="http://www.signavio.com">www.signavio.com</a>	Yes
Skand AI	Skand	<a href="http://www.skand.ai">www.skand.ai</a>	No
Software AG	Aris	<a href="http://www.arismining.com">www.arismining.com</a>	Yes
Soroco	Scout Platform	<a href="http://www.soroco.com">www.soroco.com</a>	No
StereoLogic	StereoLogic Process Mining	<a href="http://www.stereologic.com">www.stereologic.com</a>	No
TU/e	ProM	<a href="http://www.promtools.org">www.promtools.org</a>	Yes
TU/e	RapidProM	<a href="http://www.rapidprom.org">www.rapidprom.org</a>	Yes
UI Path	UI Path Process Mining	<a href="http://www.uipath.com">www.uipath.com</a>	Yes
UltimateSuite	UltimateSuite TM/RPA	<a href="http://www.ultimatesuite.com">www.ultimatesuite.com</a>	No
Upflux	Upflux	<a href="http://www.upflux.net">www.upflux.net</a>	No
Worksoft	Worksoft	<a href="http://www.worksoft.com">www.worksoft.com</a>	No

## 4.1 Pm4py

Η pm4py είναι μια βιβλιοθήκη της Python που υποστηρίζει αλγόριθμους εξόρυξης διαδικασιών (state-of-the-art), είναι ανοικτού κώδικα (με άδεια χρήσης GPL<sup>1</sup>) και προορίζεται να χρησιμοποιείται τόσο σε ακαδημαϊκά όσο και σε παραγωγικά έργα. Έχει δημιουργηθεί και αναπτύσσεται από την Fraunhofer Institute for Applied Information Technology <sup>2</sup> και πρόκειται για μια ισχυρή βιβλιοθήκη που παρέχει μια ολοκληρωμένη λύση για ανάλυση με χρήση της γλώσσας προγραμματισμού Python. Με τη χρήση της *pm4Py*, μπορεί να γίνει ανακάλυψη, παρακολούθηση και βελτίωση των επιχειρησιακών διαδικασιών με βάση τα πραγματικά δεδομένα. Τα κύρια χαρακτηριστικά που προσφέρει περιγράφονται παρακάτω.

Αρχικά προσφέρει μεθόδους "εισαγωγής αρχείου καταγραφής γεγονότων" από διάφορες μορφές αρχείων, όπως XES, CSV και ProM logs. Επιπλέον προσφέρει μεθόδους "προ-επεξεργασίας του αρχείου καταγραφής συμβάντων" και της δομής που προκύπτει (dataframe, event log). Μέθοδοι συμπεριλαμβανομένου του φιλτραρίσματος, της μετονομασίας των ιδιοτήτων, της ευθυγράμμισης χρονικών σημείων και τον χειρισμό λάθους τιμών ιδιοτήτων. Επίσης σε συνδυασμό με άλλες βιβλιοθήκες της Python, μπορεί εύκολα να επεξεργαστεί τα δεδομένα και να δημιουργήσει την ιδανική είσοδο για τους αλγόριθμους εξόρυξης. Επίσης για το στάδιο της "Ανακάλυψης Διαδικασίας" παρέχει μια ευρεία γκάμα αλγορίθμων ανακάλυψης διαδικασίας για την αυτόματη δημιουργία μοντέλων διαδικασίας από δεδομένα καταγραφής συμβάντων. Οι αλγόριθμοι αυτοί περιλαμβάνουν τον αλγόριθμο-*a*, τον inductive miner, τον heuristic miner, τον fuzzy miner και πολλούς άλλους. Για το στάδιο του "Έλεγχου Συμμόρφωσης" παρέχει τρεις μεθόδους (token replay, alignments, footprints) για την σύγκριση των ανακαλυφθέντων μοντέλων διαδικασίας με την παρατηρούμενη συμπεριφορά και ώστε να εκτελεί έλεγχο συμμόρφωσης για την ανίχνευση παραβάσεων ή ατασθαλιών στις διαδικασίες. Τέλο, παρέχει εργαλεία για την οπτικοποίηση των δεδομένων καταγραφής συμβάντων και των ανακαλυφθέντων μοντέλων διαδικασίας, με σκοπό την ανάλυση και την αντίληψη των διαδικασιών.

Τα πλεονεκτήματα χρήσης της pm4py είναι ότι πρόκειται για μια ισχυρή και πλήρης βιβλιοθήκη ανάλυσης διαδικασιών που προσφέρει μια ευρεία γκάμα αλγορίθμων, εργαλείων και λειτουργιών για την ανάλυση διαδικασιών. Παρέχει εκτεταμένη υποστήριξη για εισαγωγή δεδομένων, προ-επεξεργασία, ανακάλυψη διαδικασιών, έλεγχο συμμόρφωσης και οπτικοποίηση. Επιπλέον, ενσωματώνεται σε έργα λογισμικού που χρησιμοποιούν την Python η οποία είναι μια δημοφιλής γλώσσα για την απλότητά της και την ευελιξία της και η δυνατότητα χρήσης της στην Python επιτρέπει την αρκετά καλή ενσωμάτωση με άλλες βιβλιοθήκες ανάλυσης δεδομένων και μηχανικής μάθησης, επιτρέποντας πιο προηγμένες διαδικασίες ανάλυσης διαδικασιών. Επιπλέον, προσφέρει ένα ευέλικτο πλαίσιο που επιτρέπει στους χρήστες να προσαρμόσουν και να επεκτείνουν τις λειτουργίες της. Παρέχει διάφορες επιλογές διαμόρφωσης και παραμέτρων ώστε να μπορεί να προσαρμοστεί η συμπεριφορά των αλγορίθμων για να λειτουργήσουν σε συγκεκριμένες περιπτώσεις χρήσης ή σε συγκεκριμένα χαρακτηριστικά δεδομένων.

Στα μειονεκτήματα, είναι ότι έχει περιορισμένη απόδοση και η εκτέλεση αλγορίθμων εξόρυξης είναι πολύ χρονοβόρα, ειδικά όταν επεξεργάζεται μεγάλα σύνολα δεδομένων ή πολύπλοκες δομές διαδικασίας γεγονός που μπορεί να περιορίσει την απόδοση και την κλιμάκωση της ανάλυσης. Επιπλέον, δεν μπορεί να υποστηρίξει πλήρως τον θόρυβο και

<sup>1</sup><https://www.gnu.org/licenses/gpl-3.0.en.html>

<sup>2</sup><https://www.fit.fraunhofer.de/en.html>



τα ατελή δεδομένα που υπάρχουν στο αρχείο καταγραφής γεγονότων με αποτέλεσμα η ανάλυση να οδηγήσει σε ανακατασκευή μοντέλων που δεν αντιπροσωπεύουν πλήρως τις πραγματικές διαδικασίες. Επίσης, δεν μπορεί να διαχειριστεί πλήρως πολύπλοκες δομές με αποτέλεσμα, όταν ανακαλύπτει και αναλύει διαδικασίες με πολλαπλές διακλαδώσεις και συγχωνεύσεις, αλληλεπιδράσεις παράλληλων διαδικασιών και πολλαπλά επίπεδα ενσωμάτωσης να παρουσιάσει δυσνόητα μοντέλα. Τέλος, η χρήση της *pm4py* απαιτεί κάποιο επίπεδο εξοικείωσης με τη γλώσσα προγραμματισμού Python με αποτέλεσμα να είναι αποτρεπτικό για άτομα που δεν έχουν δεξιότητες προγραμματισμού.

## 4.2 Disco

Το Disco είναι ένα αρκετά εύχρηστο εργαλείο εξόρυξης διαδικασιών που αναπτύχθηκε από τη Fluxicon<sup>3</sup>. Παρέχει ένα ισχυρό και ευανάγνωστο περιβάλλον, αρκετά φιλικό προς το χρήστη, για την ανάλυση αρχείων καταγραφής γεγονότων και την ανακάλυψη διαδικασιών με πληθώρα τεχνικών και μεθόδων. Το Disco είναι ένα εμπορικό λογισμικό το οποίο μπορεί να χρησιμοποιηθεί τόσο για ακαδημαϊκή έρευνα, καθώς και καθιστά εύκολη την ανάπτυξη νέων αλγόριθμων από τους ερευνητές, όσο και για επαγγελματική χρήση από ανθρώπους που δεν έχουν το αντίστοιχο υπόβαθρο.

Το Disco υποστηρίζει διάφορες μορφές αρχείων, όπως XES και CSV, καθιστώντας εύκολη την εισαγωγή δεδομένων συμβάντων από διάφορες πηγές, επίσης υποστηρίζει όλους τους προηγμένους αλγόριθμους εξόρυξης διαδικασιών για την κατασκευή μοντέλων διαδικασίας, όπως τα BPMN, τα PetriNet και τα DFGs. Επιπλέον, υποστηρίζει τεχνικές ελέγχου συμμόρφωσης επισημαίνοντας αποκλίσεις, αδιέξοδα, προβλήματα μη συμμόρφωσης και παρέχει λεπτομερείς μετρικές απόδοσης, όπως ο χρόνος κύκλου διαδικασίας και η αξιοποίηση των πόρων, βοηθώντας στον εντοπισμό περιοχών για βελτίωση. Τέλος, το σημαντικό χαρακτηριστικό και αρκετά βοηθητικό για την εύλικτη εξερεύνηση και ανάλυση δεδομένων, είναι η δυνατότητα για οπτικοποίηση και αναπαράσταση της διαδικασίας με την χρήση μοντέλων, αρχείων καταγραφής και μετρικών απόδοσης με σκοπό την διευκόλυνση του αναλυτή που την εξετάζει.

Το μεγάλο πλεονέκτημα του εργαλείου είναι το ευχάριστο και το φιλικό προς τον χρήστη User Interface που το καθιστά προσβάσιμο για επαγγελματίες ανεξάρτητα του τεχνολογικού τους υπόβαθρου. Στα πλεονεκτήματα συγκαταλέγεται η ταχύτητα που επιτρέπει τη γρήγορη ανάλυση των δεδομένων και άμεση ανακάλυψη των διαδικασιών και την κλιμάκωση, ώστε να χειρίζεται αποτελεσματικά μεγάλα σε μέγεθος αρχεία, επιτρέποντας την ανάλυση πολύπλοκων διαδικασιών με πολλές διακλαδώσεις.

## 4.3 ProM

Το ProM, αντίθετα από το Disco, είναι ένα εργαλείο ανοικτού κώδικα που χρησιμοποιείται για την εξόρυξη και ανάλυση διαδικασιών. Αναπτύσσεται από το Τεχνολογικό Πανεπιστήμιο του Άντχουβεν<sup>4</sup> στην Ολλανδία και είναι διαθέσιμο για κατέβασμα σε οποιόν θέλει να εργαστεί και να πειραματιστεί με την εξόρυξη διαδικασιών<sup>5</sup>. Παρέχει ένα ολοκληρωμένο γραφικό περιβάλλον για την έρευνα, την ανάπτυξη και την εφαρμογή ενός έργου εξόρυξης διαδικασιών και προσφέρει ένα σύνολο εργαλείων και λει-

<sup>3</sup><https://fluxicon.com/>

<sup>4</sup><https://www.tue.nl/en/>

<sup>5</sup><https://promtools.org/>

τουργιών κατατάσσοντάς το σε ένα αξιόπιστο και προσβάσιμο εργαλείο ακόμα και για αρχαίους χρήστες. Αναλύει και οπτικοποιεί τις διαδικασίες, παρέχει όλες τις τεχνικές που έχουν τα υπόλοιπα εργαλεία που αναφέρθηκαν, όπως η ανακάλυψη διαδικασίας, ο έλεγχος συμμόρφωσης, η ανάλυση κοινωνικών δικτύων, η οργανωτική εξόρυξη, το clustering, η εξόρυξη αποφάσεων, η πρόβλεψη, η αναπαράσταση και η οπτικοποίηση της διαδικασίας.

Μπορεί να χρησιμοποιηθεί για την βελτίωση των διαδικασιών σε επίπεδο απόδοσης και ποιότητας υπηρεσιών και τον εντοπισμό ανωμαλιών ή αποκλίσεων, που μπορεί να οδηγήσουν σε προβλήματα. Πλεονεκτήματα, είναι το γραφικό περιβάλλον το οποίο το κάνει προσιτό στον μέσο χρήστη, η ταχύτητα του ως προς την ανακάλυψη διαδικασιών και την ανάλυση των δεδομένων και οι τεχνικές φιλτραρίσματος των περιεχομένων του αρχείου καταγραφής δεδομένων και συγχώνευσης τεχνικών με σκοπό την εξαγωγή συνθετών αποτελεσμάτων.



# Διαδικασία Χρηματοδοτήσεων Κοινής Αγροτικής Πολιτικής

Η Ευρωπαϊκή Ένωση δαπανά ένα μεγάλο μέρος του προϋπολογισμού της στην Κοινή Αγροτική Πολιτική (ΚΑΠ - CAP)[8]. Ανάμεσα στις δαπάνες αυτές είναι οι άμεσες πληρωμές, οι οποίες κυρίως στοχεύουν στην παροχή βασικού εισοδήματος στους αγρότες ανεξάρτητα από την παραγωγή. Το υπόλοιπο του προϋπολογισμού της ΚΑΠ προορίζεται για δαπάνες σχετικές με την αγορά και την αγροτική ανάπτυξη. Οι διαδικασίες που διέπουν τη διανομή αυτών των κεφαλαίων υπόκεινται σε πολύπλοκους κανονισμούς που καταγράφονται στους νόμους της ΕΕ και των εθνικών κρατών. Τα κράτη μέλη υποχρεούνται να λειτουργούν ένα Ολοκληρωμένο Σύστημα Διοίκησης και Ελέγχου (IACS), το οποίο περιλαμβάνει συστήματα Πληροφορικής για την υποστήριξη των πολύπλοκων διαδικασιών κατανομής επιδοτήσεων[7]. Η διαδικασία που εξετάζεται και έχει καταγραφεί στο σύνολο δεδομένων αφορά την επεξεργασία αιτήσεων για άμεσες πληρωμές της ΕΕ για τους γεωργούς της Γερμανίας από το Ευρωπαϊκό Ταμείο Αγροτικής Εγγύησης. Η διαδικασία επαναλαμβάνεται κάθε χρόνο με μικρές διαφορές λόγω των αλλαγών στους κανονισμούς της ΕΕ. [25]

## 5.1 Dataset

Το σύνολο δεδομένων εξάγεται από τα συστήματα της εταιρίας "data experts"[5] που εδρεύει στο Neubrandenburg της Γερμανίας. Το εργαλείο τους profil c/s υποστηρίζει αυτές τις διαδικασίες στο επίπεδο των ομοσπονδιακών υπουργείων γεωργίας και των τοπικών τμημάτων. Το συνολικό αρχείο καταγραφής περιέχει 43.809 αιτήσεις απευθείας πληρωμών τριετούς διάρκειάς από το 2015 έως το 2017[9]. Η συντομότερη περίπτωση περιέχει 24 γεγονότα, η μακρύτερη 2.973 και κατά μέσο όρο υπάρχουν 57 γεγονότα ανά περίπτωση που αναφέρονται σε 14 δραστηριότητες. Το σύνολο των γεγονότων ανέρχεται σε 2.514.266 και αντιπροσωπεύουν συνδυασμό αυτόματων και χειρωνακτικών διαδικασιών που περιλαμβάνουν την παραλαβή και αποδοχή της αίτησης, με στόχο την ολοκλήρωση της δηλαδή την έγκριση της πληρωμής. Τα workflows στο profil c/s μπορούν να κατανοηθούν σε όρους εγγράφων[πίνακας5.1], όπου κάθε έγγραφο έχει μια κατάσταση που επιτρέπει συγκεκριμένες ενέργειες. Αυτές οι ενέργειες μπορούν να εκτελούνται χειροκίνητα οποιαδήποτε στιγμή μέσω εργαλείων ειδικά σχεδιασμένων για το έγγραφο ή μπορούν να προγραμματιστούν αυτόματα. Τα έγγραφα αναφέρονται, είτε ρητά στο αρχείο καταγραφής είτε παρουσιάζονται σαφώς αν ένας μεγάλος αριθμός ενεργειών εκτελείται από τον ίδιο χρήστη περίπου την ίδια στιγμή. [25]. Μερικές αιτήσεις επαναλαμβάνονται από συγκεκριμένα resources και για συγκεκριμένους λόγους, είτε από το τμήμα (υποδιαδικασία "Αλλαγή"), είτε λόγω νομικής αντιρρήσεως από τον

αιτούντα (υποδιαδικασία "Αντίρρηση"), ενώ κάποιες άλλες υπόκεινται σε επιθεωρήσεις. Τέλος, κατά την διάρκεια της διαδικασίας οι αιτήσεις περνούν μέσα από αρκετά βήματα που καθορίζουν αν πρόκειται για απευθείας ή όχι πληρωμή και το ποσό που θα καταβληθεί στο τέλος στον αιτούντα. Το αν θα καταβληθεί όλο το ποσό ή ένα μέρος αυτού μπορεί να συμβεί για διάφορους λόγους, για παράδειγμα, εάν το δηλωμένο μέγεθος ενός αγροτεμαχίου δεν ταιριάζει με το πραγματικό μέγεθος, όπως προσδιορίζεται από την εξ αποστάσεως ή επιτόπου επιθεώρηση ή για λόγους που περιλαμβάνουν την μη συμμόρφωση με την αγροτική πολιτική ή τη μη συμμόρφωση με τις προϋποθέσεις κάποιου από τα χαρακτηριστικά που συνάδουν με τον αιτούντα, για παράδειγμα ένας νεαρός αγρότης έχει άλλες προϋποθέσεις από έναν γηραιότερο.

Table 5.1: Περιγραφή Εγγράφων διαδικασίας που καθορίζουν την ροή της διαδικασίας, όπου κάθε έγγραφο έχει μία κατάσταση(διαδικασία) που επιτρέπει συγκεκριμένες δραστηριότητες.

Νο	Όνομα Εγγράφου	Υποδιαδικασία	Περιγραφή
1	Control summary	Main	Έγγραφο που περιέχει τα περιληπτικά αποτελέσματα διάφορων ελέγχων (συντονισμός αναφοράς, έλεγχος τμήματος, ελέγχους)
2	Department control parcels (before 2017)	Main	Ένα έγγραφο που περιέχει τα αποτελέσματα ελέγχων σχετικά με την πιστοποίηση των αγροτεμαχίων που έχει στην κατοχή του ο αιτών.
3	Entitlement application	Main Application Objection	Η αίτηση για επιδότηση, δηλαδή το δικαίωμα να υποβληθεί αίτηση για άμεσες πληρωμές, συνήθως δημιουργείται μία φορά στην αρχή μιας νέας περιόδου χρηματοδότησης.
4	Parcel Document (before 2016)	Main	Έγγραφο που περιέχει όλα τα αγροτεμάχια για τα οποία ζητούνται επιδοτήσεις.
5	Inspection	On-Site Remote	Έγγραφο που περιέχει τα αποτελέσματα επιτόπιων ή απομακρυσμένων ελέγχων.
6	Geo Parcel Document (replaces Parcel document since 2016 and Department control parcels since 2017)	Main Declared Reported	Το έγγραφο που περιέχει όλα τα αγροτεμάχια για τα οποία ζητούνται επιδοτήσεις. Από το 2017, το Έγγραφο Γεω-Αγροτεμαχίων αντικαθιστά επίσης το έγγραφο του Ελέγχου Τμήματος.
7	Payment application	Main Application Objection Change	Αίτηση για άμεσες πληρωμές διάρκειας ενός χρόνου.
8	Reference alignment	Main	Έγγραφο που περιέχει τα τοπογραφικά των αγροτεμαχίων όπως αναφέρονται από τον αιτούντα και τα συγκρίνει με τα δηλωμένα αγροτεμάχια σε κάποια επίσημη αρχή(π.χ. κτηματολόγιο).

## 5.2 Προκλήσεις Διαδικασίας

Οι διαδικασίες χρηματοδοτήσεων της Ευρωπαϊκής Ένωσης ανεξαρτήτου τομέα (Οικονομικής Ανάπτυξης, Αγροτικής Ανάπτυξης, Μεταναστευτικού, κα.) είναι αρκετά πολύπλοκες και χρειάζεται αρκετή εμπειρία και αλληλεπίδραση για να κατανοηθούν πλήρως. Πρόκειται για διαδικασίες που διέπονται από πολύπλοκους κανονισμούς που θέτει η Ευρωπαϊκή Ένωση και ταυτόχρονα πρέπει να εναρμονίζονται με τους νόμους κάθε κράτους-μέλους της. Η συγκεκριμένη διαδικασία δεν διαφέρει από την ανωτέρω περιγραφή και πρέπει να εκτελεστεί με τέτοιο τρόπο, ώστε να ακολουθεί τους κανονισμούς που έχουν συμφωνηθεί με σκοπό να οδηγήσουν στην έγκριση ή απόρριψη της αίτησης και την τελική καταβολή ή όχι μέρος ή όλου του πόσου αποζημίωσης που δικαιούται ο αγρότης.

Πέραν των δυσκολιών που αφορούν τους κανονισμούς η διαδικασία των άμεσων πληρωμών έχει πληθώρα δραστηριοτήτων στις οποίες πέραν των πληροφοριακών συστημάτων συμμετέχει και ο ανθρώπινος παράγοντας ο οποίος είναι απρόβλεπτος και στην χρήση ενός συστήματος. Αρκετές από τις δραστηριότητες είναι χειροκίνητες και έχουν να κάνουν με ελέγχους που αφορούν εξωτερικούς πόρους, όπως είναι οι εγκρίσεις από το κτηματολόγιο για την ορθή δήλωση των αγροτεμαχίων, οι επισκέψεις μηχανικών στο πεδίο για τις αντίστοιχες μελέτες και εκτιμήσεις σχετικά με τα αγροτεμάχια, κάνουν ακόμα πιο πολύπλοκη την διαδικασία. Πρέπει, επίσης, να αναφερθεί ότι σε τέτοιες διαδικασίες επιχορηγήσεων μέσω της Ευρωπαϊκής Ένωσης, στα κράτη μέλη υπάρχει και το outsourcing σε επίπεδο αίτησης και διεκπεραίωσης της διαδικασίας. Πολλοί αγρότες αναθέτουν την εργασία αίτησης σε κάποιο χρηματοπιστωτικό ίδρυμα ή σε κάποια εξειδικευμένη επιχείρηση που παρέχει τέτοιου είδους υπηρεσίες. Οπότε είναι λογικό να βρεθούν αρκετές παραλλαγές στην εκτέλεση της, διαφορετικά εκτελεί την διαδικασία κάποιος που την κάνει πρώτη φορά και διαφορετικά ένας που έχει εμπειρία αναφορικά με τον χειρισμό παρόμοιων συστημάτων.

Ιδιαίτερη πρόκλησή αποτελούν τα προβλήματα που προκύπτουν από την εκτέλεση της διαδικασίας και τα αποτελέσματα τα οποία δεν ικανοποιούν τους μετέχοντες σε αυτή. Φυσιολογικό αποτέλεσμα θεωρείται η διεκπεραίωση της αίτησης που κατατίθεται σε μια συγκεκριμένη χρονιά και ολοκληρώνεται μέχρι το τέλος του έτους. Αυτό συμβαίνει ίσως γιατί καθυστερεί η διαδικασία ή πρέπει να ξανανοίξει για ορισμένους λόγους, όπως η λάθος κατάθεση εγγραφών με αποτέλεσμα να υπάρχουν επιπτώσεις στο τελικό ποσό αποζημίωσης. Επίσης, σε επίπεδο διεκπεραίωσης διαδικασίας, ιδιαίτερο ρόλο παίζει το κατά πόσο τα διάφορα τμήματα μπορούν να την εκτελέσουν όσο γίνεται πανομοιότυπα και να μην υπάρχουν διάφορες μεταξύ τους. Σε γενικές γραμμές μεγάλη πρόκλησή αποτελεί η διαχείριση τις απρόβλεπτης ενέργειας του ανθρώπινου παράγοντα σε κάποια στιγμή της διαδικασίας, κάτι το οποίο μπορεί να οδηγήσει σε καθυστερήσεις και περιπτώσεις χρήσης πολύ διαφορετικές από τις αναμενόμενες.



# Ερευνητική Μεθοδολογία

## 6.1 Import Event Log

Για τους σκοπούς της παρούσας εργασίας θα γίνει χρήση των μεθόδων που παρέχει η *pm4py* για την εισαγωγή του αρχείου καταγραφής συμβάντων από διάφορες μορφές αρχείων, όπως XES, CSV και ProM logs. Το αρχείο καταγραφής δεδομένων το οποίο έχει προέλθει από το Τεχνολογικό Πανεπιστήμιο του Άιντχοβεν<sup>1</sup> σε μορφή πρότυπου IEEE XES, θα διαβαστεί και θα αποθηκευτεί με δύο μεθόδους. Αρχικά στην σταθερή μνήμη (σκληρό δίσκο) του υπολογιστή σε .csv αρχείο για χρήση εκτός του προγράμματος που έχει κατασκευαστεί και σε μορφή dataframe της Python το οποίο αποθηκεύεται στην RAM για όσο διαρκεί η λειτουργία του προγράμματος και θα αποτελέσει είσοδο σχεδόν για όλες τις συναρτήσεις που αφορούν το έργο.

## 6.2 Κατανόηση Διαδικασίας μέσω των δεδομένων

Ένα από τα πλεονεκτήματα που αναφέρθηκε και στην παρουσίαση του προτύπου IEEE XES είναι το γεγονός ότι μπορεί να αποθηκεύσει πολλές πληροφορίες μέσω των ιδιοτήτων της διαδικασίας, είτε σε επίπεδο διαδικασίας είτε σε επίπεδο γεγονότων. Ιδιότητες όπως είναι το *applicant* που κρατάει το id του αιτούντος και είναι ίδιο ανά τα χρονιά, το *Department*, που περιέχει το id του τμήματος που εξυπηρετεί κάθε περίπτωση και πολλές άλλες. Επωφελούμενοι των ιδιοτήτων του αρχείου καταγραφής γεγονότων, θα γίνει προσπάθεια, στο μεγαλύτερο βαθμό που είναι αυτό εφικτό, να κατανοηθεί η διαδικασία και το εννοιολογικό πλαίσιο στο οποίο λειτουργήσει. Για τον σκοπό αυτό θα χρησιμοποιηθούν εκτός από την *pm4py*, κλασσικές βιβλιοθήκες της Python που χρησιμοποιούνται συχνά στον τομέα εξόρυξης διαδικασιών, η *numpy*, η *pandas* και η *matplotlib* για την οπτικοποίηση των δεδομένων μέσω plots.

Βάσει αυτών των ιδιοτήτων και μετά την μετατροπή του αρχείου .xes σε dataframe θα γίνει μια πρώτη παρουσίαση των δεδομένων, όπως ο συνολικός αριθμός των περιπτώσεων και γεγονότων, την κατανομή ανά χρονιά και μήνα και ο αριθμός αιτήσεων που εξυπηρετεί το κάθε τμήμα. Αυτές και παρόμοιες πληροφορίες με την χρήση πινάκων και διαγραμμάτων θα δώσουν μια αρχική εικόνα για τον τρόπο λειτουργίας των αιτήσεων για άμεσες πληρωμές μέσω του προϋπολογισμού της Ευρωπαϊκής Ένωσης για την Κοινή Αγροτική Πολιτική.

---

<sup>1</sup><https://www.win.tue.nl/bpi/2018/challenge.html>



### 6.3 Process Discovery

Όπως αναφέρθηκε και στην θεωρία υπάρχουν αρκετές τεχνικές και αλγόριθμοι διαδικασιών οι οποίοι έχουν σαν στόχο την εξαγωγή του βέλτιστου μοντέλου διαδικασιών. Η *pm4py* προσφέρει σχεδόν όλους τους διαθέσιμους αλγόριθμους εξόρυξης διαδικασιών και όλες τις διαθέσιμες σημειογραφίες και μοντέλα οπτικοποίησης διαδικασιών. Επίσης τα συγκεκριμένα API μπορούν και παραμετροποιούνται σύμφωνα με την θεωρία και την λογική της κάθε τεχνοτροπίας.

Στην παρούσα ανάλυση, αρχικά για να μπορέσει να γίνει εξαγωγή ενός πιο κατανοητού μοντέλου από κάποιον που δεν είναι εξοικειωμένος με την τεχνολογία της εξόρυξης διαδικασιών και την ανάγνωση των μοντέλων, έγινε μια προ-επεξεργασία στα δεδομένα έτσι ώστε η διαδικασία να παρουσιαστεί σε υψηλό επίπεδο, δηλαδή σε επίπεδο εγγράφου. Στην συνέχεια θα παρουσιαστούν διαφορετικά μοντέλα και οι ιδιότητές τους καθώς και μία πρώτη αξιολόγηση τους σε υψηλό επίπεδο ανάλογα με τα χαρακτηριστικά τους. Συγκεκριμένα θα εφαρμοστούν οι τρεις αλγόριθμοι εξόρυξης διαδικασιών που αναφέρθηκαν στο θεωρητικό μέρος και θα παράξουν σύμφωνα με τα χαρακτηριστικά τους τα αντίστοιχα μοντέλα. Για παράδειγμα ο αλγόριθμος-*a* παράγει ένα PetriNet και ο Inductive Miner ένα δένδρο διαδικασιών το οποίο στην συνέχεια θα μετατραπεί σε PetriNet και BPMN. Το μοντέλο το οποίο εξάγεται αποτελεί είσοδο για τα υπόλοιπα στάδια ανάλυσης της διαδικασίας όπως ο έλεγχος συμμόρφωσής, η αξιολόγηση της διαδικασίας, η βελτίωση και η ανακάλυψη των κοινωνικών δικτύων.

### 6.4 Conformance Checking

Μετά την ανακάλυψη ενός συνόλου μοντέλων διαδικασιών με την χρήση των βασικών αλγορίθμων εξόρυξης διαδικασιών και τεχνικών που βασίζονται σε αυτούς και υποστηρίζονται από τη βιβλιοθήκη *pm4py* το επόμενο βήμα είναι ο έλεγχος συμμόρφωσής. Με τον έλεγχο συμμόρφωσης θα γίνει σύγκριση μεταξύ των μοντέλων διαδικασίας που έχουν ανακαλυφθεί και της παρατηρούμενης συμπεριφοράς που έχει καταγραφεί και αξιολόγηση βάσει των κριτηρίων που έχουν αναφερθεί στην βιβλιογραφία. Στόχος είναι να ελεγχθεί εάν το αρχείο καταγραφής συμμορφώνεται με το μοντέλο και αντίστροφα. Επίσης στο πλαίσιο του ελέγχου συμμόρφωσης θα μετρηθούν και θα αξιολογηθούν οι αλγόριθμοι εξόρυξης διαδικασιών.

Με τον έλεγχο συμμόρφωσης θα γίνει εφαρμογή των διαφορών τεχνικών ώστε οι αποκλίσεις που θα ανακαλυφθούν να αναλυθούν από δύο οπτικές, την περίπτωση να υπάρχει λάθος στο μοντέλο ή την περίπτωση της λάθος εκτέλεσης της διαδικασίας. Τα κύρια ερωτήματα που δημιουργούνται είναι, αν το μοντέλο είναι λάθος και δεν ανταποκρίνεται στην πραγματικότητα, πως βελτιώνεται; Αν η διαδικασία είναι λάθος και ορισμένες περιπτώσεις αποκλίνουν από το μοντέλο και απαιτούν διορθωτικές κινήσεις, πως βελτιώνεται η ροή της διαδικασίας ώστε να πετύχουμε καλύτερη συμμόρφωσή;

Στην βιβλιοθήκη *pm4py*, υλοποιούνται δύο βασικές τεχνικές: Token-based replay και των ευθυγραμμίσεων (alignments). Προς το παρόν, υποστηρίζει επιπλέον και τεχνικές με την χρήση FootPrint, αλλά σε επόμενες εκδόσεις της θα καταργηθεί. Έχουν δημιουργηθεί συναρτήσεις οι οποίες εκμεταλλεύονται πλήρως τις τεχνικές που παρέχει η βιβλιοθήκη και εκτελεί εργασίες ελέγχου συμμόρφωσής. Σαν είσοδο έχει το ανακαλυφθέν μοντέλο και την παρατηρούμενη συμπεριφορά όπως αυτή έχει καταγραφεί στα event logs και παρέχει τα αντίστοιχα αποτελέσματα σε μορφή διαγνωστικών.

Επίσης, σε όποιες τεχνικές αυτό είναι δυνατό, παρουσιάζεται ένα δείγμα σύγκρισης μεταξύ μοντέλων και περιπτώσεων χρήσης και των αντίστοιχων αποτελεσμάτων τους. Για την τεχνική Token-Replay για μια περίπτωση χρήσης παρουσιάζει αν το συγκεκριμένο ίχνος ταιριάζει και τον αριθμό των tokens που λείπουν, έχουν παραχθεί, καταναλωθεί ή μένουν ενεργά στο PetriNet. Για την τεχνική των ευθυγραμμίσεων παρουσιάζει το κόστος της ευθυγράμμισης, το αν ταιριάζει ή όχι και μια σειρά στατιστικών για τον αριθμό των βημάτων που χρειάστηκε ή τον αριθμό των καταστάσεων που χρησιμοποίησε. Ενδεικτικά, παρουσιάζεται και ένα σύνολο ευθυγραμμίσεων έτσι ώστε να γίνει κατανοητή η συγκεκριμένη τεχνική. Τέλος, για τα footprints παρουσιάζονται αντίστοιχα διαγνωστικά, εικόνες των αποτελεσμάτων του υπολογισμού του footprint που αναπαριστά το μοντέλο και σειρά από στατιστικά τα οποία χρησιμεύουν στην εξαγωγή συμπερασμάτων.

## 6.5 Αξιολόγηση Αλγόριθμων Εξόρυξης Διαδικασιών

Σε αυτό το στάδιο του έργου θα αξιολογήσουμε με την χρήση της βιβλιοθήκης της python *pm4py* τα αποτελέσματα του σταδίου της "Ανακάλυψης Διαδικασιών", δηλαδή των μοντέλων διαδικασιών που ανακαλύφθηκαν ως προς τα τέσσερα κριτήριά όπως αυτά παρουσιάστηκαν στο βιβλιογραφικό μέρος. Θα αξιολογηθεί κάθε μοντέλο που ανακαλύφθηκε ξεχωριστά σε αντιπαράθεση με το αρχείο καταγραφής σύμφωνα με τη θεωρία και την χρήση των συναρτήσεων που κατασκευάστηκαν για να κάνουν τους αντίστοιχους υπολογισμούς. Για να εξοικονομηθεί υπολογιστικός χρόνος πολλοί από τους υπολογισμούς, όπως ο υπολογισμός του *fitness* και του *precision* εντάχθηκαν στα πλαίσια του ελέγχου συμμόρφωσης και για αυτό θα παρατηρηθεί ότι η συνολική καταλληλότητά του αλγόριθμου επηρεάζεται από την εκάστοτε τεχνική. Αναφέρθηκε στην βιβλιογραφία ο διαφορετικός τρόπος που η κάθε μέθοδος επηρεάζει τις τιμές υπολογισμού των κριτηρίων αξιολόγησης.

## 6.6 Ανάλυση Διαδικασίας

Η ανάλυση της διαδικασίας θα επικεντρωθεί στη σχέση μεταξύ ανθρώπων, πληροφοριακών συστημάτων και διαδικασιών, πρόκειται για δεδομένα που αφορούν την κοινωνική πτυχή της ανάλυσης. Σκοπός είναι αρχικά η ανακάλυψη των πόρων (resources) που συμμετέχουν στην διαδικασία και η δημιουργία πινάκων εργασίας και δραστηριοτήτων. Θα γίνει ανακάλυψη ενός κοινωνικού δικτύου από τα δεδομένα της διαδικασίας σε επίπεδο δραστηριοτήτων (event data) τα οποία έχουν την δυνατότητα να αποθηκεύουν την ανθρώπινη συμπεριφορά και περιέχουν πληροφορίες για το ποιος έχει πραγματικά εκτελέσει την διαδικασία. Θα κατασκευαστεί ο *Resource Activity Matrix* με τον οποίο ανακαλύπτεται ο χρόνος που σπατάλησε ένα resource σε μια δραστηριότητα. Στην συνέχεια το Κοινωνικό δίκτυο (Social NetWork) της διαδικασίας όπου έρχεται να συμπληρώσει τον πίνακα δραστηριοτήτων και να παρουσιάσει τις σχέσεις συνεργασίας των εργαζομένων. Θα γίνουν μετρήσεις όπως ο βαθμός κεντρικότητας, ο συνολικός βαθμός του δικτύου, το closeness centrality, το betweenness centrality και άλλοι υπολογισμοί που βασίζονται στην θεωρία γραφημάτων. Στην συνέχεια, θα παρουσιαστεί στον μέγιστο δυνατό βαθμό, εξαιτίας της πολυπλοκότητας και του μεγέθους των δεδομένων, μία decision point ανάλυση βάσει ενός κοινού χαρακτηριστικού των περιπτώσεων της διαδικασίας.



# Τεχνική Υλοποίηση

## 7.1 Λειτουργικές Απαιτήσεις

Για τις ανάγκες της εργασίας δημιουργήθηκε ένας πηγαίος κώδικας σε μορφή Python Script ο οποίος έχει σαν είσοδο ένα αρχείο καταγραφής γεγονότων σε μορφή προτύπου IEEE XES και το οποίο, στην συνέχεια, με συγκεκριμένη τεχνοτροπία μετατρέπεται σε δομή δεδομένων dataframe για να χρησιμοποιηθεί στην συνέχεια από τις υπόλοιπες συναρτήσεις. Χρησιμοποιήθηκε η έκδοση της Python 3.11 σε λειτουργικό σύστημα Windows 10 και παρακάτω παρουσιάζονται οι βασικές βιβλιοθήκες που χρειάζεται να εγκατασταθούν, έτσι ώστε ο Python Interpreter να εκτελέσει σωστά το script.

Table 7.1: Βασικές Βιβλιοθήκες της Python που χρησιμοποιήθηκαν για στην κατασκευή του Script.

Library	Version
pm4py	2.7.4
prettytable	3.7.0
pip	21.1.2
pandas	2.0.2
openai	0.27.7
numpy	1.24.3
matplotlib	3.7.1
graphviz	0.20.1

## 7.2 Περιγραφή Συναρτήσεων

Παρακάτω αναλύονται οι συναρτήσεις που δημιουργήθηκαν με την γλώσσα προγραμματισμού Python για να μπορέσει να γίνει καλύτερη ανάλυση της διαδικασίας. Οι συναρτήσεις χωρίζονται σε αυτές που δουλεύουν γενικά σε όλα τα dataset που χρησιμοποιούν το πρότυπο IEEE XES και σε αυτές που έχουν εξατομικευτεί για το dataset το οποίο αναλύει το συγκεκριμένο κείμενο.

### 7.2.1 Γενικές Συναρτήσεις

#### TRACE()

---

<b>Description</b>	Εκτυπώνει το μήνυμα το οποίο έχει σαν όρισμα στο αρχείο που κρατάει τα logs που αφορούν την λειτουργία του προγράμματος.
<b>Prototype</b>	TRACE(trace).
<b>Parameters</b>	trace:str
<b>Return Values</b>	-

#### TRACE1()

---

<b>Description</b>	Εκτυπώνει το μήνυμα και την μεταβλητή τα οποία έχουν σαν ορίσματα στο αρχείο που κρατάει τα logs που αφορούν την λειτουργία του προγράμματος.
<b>Prototype</b>	TRACE(trace, value).
<b>Parameters</b>	trace:str, value
<b>Return Values</b>	-

#### TRACEb()

---

<b>Description</b>	Εκτυπώνει το μήνυμα και την μεταβλητή ,σε μορφή bytes, τα οποία έχει σαν ορίσματα στο αρχείο που κρατάει τα logs που αφορούν την λειτουργία του προγράμματος.
<b>Prototype</b>	TRACEb(trace, value).
<b>Parameters</b>	trace:str, value
<b>Return Values</b>	-

#### print\_logs\_from\_xml()

---

<b>Description</b>	Αποθηκεύει και εκτυπώνει την δομή και τις βασικές ιδιότητες του xml αρχείου που περιέχει τα αρχεία καταγραφής. Χρησιμοποιείται για λόγους debugging σε περίπτωση που κριθεί αναγκαίο εξαιτίας του μεγάλου μεγέθους του.
<b>Prototype</b>	print_logs_from_xml(dataset:str).
<b>Parameters</b>	dataset:str
<b>Return Values</b>	-

## ask\_chatGPT

---

<b>Description</b>	Με την χρήση της βιβλιοθήκης <i>pm4py</i> κατασκευάζει καταλληλο ερώτημα και συνδέεται με την OpenAi η οποία θα δώσει την αντίστοιχη απάντηση που εκτυπώνεται στο αντίστοιχο αρχείο.
<b>Prototype</b>	<code>ask_chatGPT(query:str).</code>
<b>Parameters</b>	<code>query:str</code>
<b>Return Values</b>	-

## 7.2.2 Εξαγωγή - εισαγωγή δεδομένων

### import\_xes\_files()

---

<b>Description</b>	Δέχεται σαν όρισμα ένα .xes αρχείο και επιστρέφει το event log σε μορφή Dataframe.
<b>Prototype</b>	<code>import_xes_files(filename).</code>
<b>Parameters</b>	<code>filename:str</code>
<b>Return Values</b>	<code>log:Dataframe</code>

### export\_from\_xes\_to\_csv()

---

<b>Description</b>	Μετατρέπει και αποθηκεύει στο μονοπάτι που έχει οριστεί σαν όρισμα ένα Dataframe σε μορφή .csv αρχείου.
<b>Prototype</b>	<code>export_from_xes_to_csv(filename, log).</code>
<b>Parameters</b>	<code>filename:str, event_log:Dataframe</code>
<b>Return Values</b>	-

## 7.2.3 Στατιστικά

### data\_number()

---

<b>Description</b>	Υπολογίζει και επιστρέφει το συνολικό αριθμό των events και των traces που υπάρχουν μέσα στο event log που έχει δεχθεί σαν όρισμα.
<b>Prototype</b>	<code>data_number(event_log).</code>
<b>Parameters</b>	<code>event_log:Dataframe</code>
<b>Return Values</b>	<code>total_traces:int, total_events:int</code>

**number\_of\_events()**


---

<b>Description</b>	Υπολογίζει και επιστρέφει το συνολικό αριθμό των events.
<b>Prototype</b>	<code>number_of_events(event_log).</code>
<b>Parameters</b>	<code>event_log:Dataframe</code>
<b>Return Values</b>	<code>len(event_stream):int</code>

**throughput\_time()**


---

<b>Description</b>	Υπολογίζει για το αρχείο ή τα αρχεία καταγραφής δεδομένων που έχει σαν είσοδο τον μέσο όρο διεκπεραίωσης μιας περίπτωσης διαδικασίας, ανακαλύπτει την μικρότερη και μεγαλύτερη σε χρόνο περίπτωση, στην συνέχεια κατασκευάζει το αντίστοιχο plot με τα αποτελέσματα τα οποία υπολόγισε.
<b>Prototype</b>	<code>throughput_time(event_log: list, filepath: str).</code>
<b>Parameters</b>	<code>event_log: list,</code> <code>filepath: str</code>
<b>Return Values</b>	-

**print\_variant\_seq()**


---

<b>Description</b>	Υπολογίζει και εκτυπώνει τον αριθμό των διαφορετικών περιπτώσεων (variants) και πόσες από αυτές είναι μοναδικές μέσα σε ένα αρχείο καταγραφής γεγονότων.
<b>Prototype</b>	<code>print_variant_seq(variants).</code>
<b>Parameters</b>	<code>variants: dict[tuple[str], list[Trace]]</code>
<b>Return Values</b>	-

**find\_same\_applicants\_by\_year()**


---

<b>Description</b>	Υπολογίζει και επιστρέφει τον αριθμό των περιπτώσεων που έχουν κοινά χαρακτηριστικά μεταξύ δύο αρχείων καταγραφής γεγονότων .
<b>Prototype</b>	<code>find_same_applicants_by_year(df_start_year_a, df_start_year_b, years: list).</code>
<b>Parameters</b>	<code>df_start_year_a:dataframe,</code> <code>df_start_year_b:dataframe,</code> <code>years:list</code>
<b>Return Values</b>	<code>len(common_values):int</code>

**find\_month\_applications()**

---

<b>Description</b>	Επιστρέφει ένα dictionary με τον αριθμό των δραστηριοτήτων εκκίνησης της διαδικασίας ανά μήνα.
<b>Prototype</b>	<code>find_month_applications(df_start_year)</code> .
<b>Parameters</b>	<code>df_start_year_a: dataframe</code>
<b>Return Values</b>	<code>month_values: dictionary</code>

**penalties\_distribution()**

---

<b>Description</b>	Επιστρέφει και εκτυπώνει ένα dictionary που περιέχει την κατανομή των ποινών που αναφέρονται στο συγκεκριμένο dataset ανά έτος.
<b>Prototype</b>	<code>penalties_distribution(start_year, year)</code> .
<b>Parameters</b>	<code>start_year: dataframe,</code> <code>year: str</code>
<b>Return Values</b>	<code>penalties: dictionary</code>

**start\_activities\_vis()**

---

<b>Description</b>	Υπολογίζει και επιστρέφει ένα dictionary με τις αρχικές δραστηριότητες και τον αριθμό συχνότητας εμφάνισης της κάθε μίας. Επίσης, τις εκτυπώνει σε μορφή plot.
<b>Prototype</b>	<code>start_activities_vis(log, filepath: str, stats: bool = True)</code> .
<b>Parameters</b>	<code>log: Dataframe,</code> <code>filepath: str,</code> <code>stats: bool = True</code>
<b>Return Values</b>	<code>start_activities: dictionary</code>

**end\_activities\_vis()**

---

<b>Description</b>	Υπολογίζει και επιστρέφει ένα dictionary με τις δραστηριότητες τερματισμού και τον αριθμό συχνότητας εμφάνισης της κάθε μίας. Επίσης, τις εκτυπώνει σε μορφή plot.
<b>Prototype</b>	<code>end_activities_vis(log, filepath: str, stats: bool = True)</code> .
<b>Parameters</b>	<code>log: Dataframe,</code> <code>filepath: str,</code> <code>stats: bool = True</code>
<b>Return Values</b>	<code>end_activities: dictionary</code>



**discover\_and\_visualize\_department()**


---

<b>Description</b>	Υπολογίζει σε ένα σύνολο από αρχεία καταγραφής γεγονότων πόσες περιπτώσεις διαδικασίας εκτελούνται ανά τμήμα
<b>Prototype</b>	<code>discover_and_visualize_department(log: list, departments: list, years: list).</code>
<b>Parameters</b>	<code>log:Dataframe,</code> <code>departments: list,</code> <code>years: list</code>
<b>Return Values</b>	-

**plot\_number\_of\_cases\_per\_year()**


---

<b>Description</b>	Βάσει ορισμάτων δημιουργεί ένα plot που παρουσιάζει πόσες περιπτώσεις της διαδικασίας έχουν εξυπηρετηθεί ανά έτος
<b>Prototype</b>	<code>plot_number_of_cases_per_year(event_logs: list, years: list, filepath: str).</code>
<b>Parameters</b>	<code>log:list,</code> <code>years: list,</code> <code>filepath:str</code>
<b>Return Values</b>	-

**count\_unique\_cases()**


---

<b>Description</b>	Ανακαλύπτει πόσες παραλλαγές εκτέλεσης της διαδικασίας (variants) υπάρχουν σε ένα αρχείο καταγραφής γεγονότων.
<b>Prototype</b>	<code>count_unique_cases(variants: dictionary).</code>
<b>Parameters</b>	<code>variants:dictionary</code>
<b>Return Values</b>	-

**7.2.4 Φιλτράρισμα****start\_year\_filtering()**


---

<b>Description</b>	Επιστρέφει event log (full, partial) βάσει της χρονιάς της πρώτης δραστηριότητας μιας περίπτωσης διαδικασίας .
<b>Prototype</b>	<code>start_year_filtering(event_log, year, activities).</code>
<b>Parameters</b>	<code>event_log:dataframe,</code> <code>year:int,</code> <code>activities:dictionary</code>
<b>Return Values</b>	<code>filtered_dataframe:dataframe,</code> <code>filtered_dataframe_year:dataframe</code>

**end\_year\_filtering()**


---

<b>Description</b>	Επιστρέφει ένα event log βάσει της χρονιάς της τελευταίας δραστηριότητας μιας περίπτωσης διαδικασίας .
<b>Prototype</b>	end_year_filtering(event_log, year, activities).
<b>Parameters</b>	event_log:dataframe, year:int, activities:dictionary
<b>Return Values</b>	filtered_dataframe:dataframe

**filter\_end\_by\_year()**


---

<b>Description</b>	Επιστρέφει ένα event log που περιέχει διαδικασίες βάσει ενός χρονικού διαστήματος σε επίπεδο έτους.
<b>Prototype</b>	filter_end_by_year(event_log_a, event_log_b, year_st, year_end).
<b>Parameters</b>	event_log_a:dataframe, event_log_b:dataframe, year_a:int, year_b:int
<b>Return Values</b>	filtered_dataframe:dataframe

## 7.2.5 Ανακάλυψη Διαδικασίας

**process\_model\_discovery()**


---

<b>Description</b>	Βάσει μεθόδου επιλέγει τον αντίστοιχο αλγόριθμο να ανακαλύψει το αντίστοιχο μοντέλο διαδικασιών. Παραμετροποιείται βάσει της θεωρίας του κάθε αλγορίθμου εξόρυξης διαδικασιών. Οι μέθοδοι ανακάλυψης που υποστηρίζει είναι "alpha miner", "alpha miner+", "inductive miner" και "heuristic miner"
<b>Prototype</b>	process_model_discovery(method: str, event_log:dataframe, noise_th: float = 0.0, dependency_th: float = 0.5).
<b>Parameters</b>	method:str, event_log:dataframe, noise_th: float = 0.0, dependency_th: float = 0.5
<b>Return Values</b>	net, initial_marking, final_marking

**bpmn\_model()**


---

<b>Description</b>	Με την χρήση του "inductive miner" ανακαλύπτεται το μοντέλο διαδικασιών σε σημειογραφία BPMN και αποθηκεύεται σε αρχείο .png
<b>Prototype</b>	bpmn_model(event_log:dataframe, filepath:str).
<b>Parameters</b>	event_log:dataframe, filepath:str
<b>Return Values</b>	-

**directly\_follows\_graphs()**


---

<b>Description</b>	Ανάλογα τις παραμέτρους κατασκευάζει έναν DFG(Directly Follows Graph).
<b>Prototype</b>	directly_follows_graph(log, filepath: str, performance: bool = False, frequency: bool = True).
<b>Parameters</b>	event_log:dataframe, filepath:str, performance: bool, frequency:bool
<b>Return Values</b>	-

**process\_tree\_inductive()**


---

<b>Description</b>	Με την χρήση του "Inductive Miner" κατασκευάζει το δένδρο διαδικασιών βάσει των παραμέτρων που έχουν δοθεί.
<b>Prototype</b>	process_tree_inductive(log, filepath:str, noise_thershold:float=0.0, multi_processing:bool=False, extract_bpmn_model:bool=False).
<b>Parameters</b>	log:dataframe, filepath:str, noise_thershold:float, multi_processing:bool, extract_bpmn_model:bool
<b>Return Values</b>	-

**extract\_petrinet()**

---

<b>Description</b>	Βάσει μεθόδου επιλέγει τον αντίστοιχο αλγόριθμο με σκοπό να ανακαλύψει το αντίστοιχο μοντέλο διαδικασίας σε μορφή PetriNet. Το εμφανίζει στο χρήστη και παράλληλα το αποθηκεύει σε .png αρχείο στο φάκελο που του επιδεικνύει ο χρήστης.
<b>Prototype</b>	<code>extract_petrinet(method: str, net, initial_marking, final_marking, filepath: str, multi_processing: bool = False).</code>
<b>Parameters</b>	<code>method:str, net, initial_marking, final_marking, filepath:str, multi_processing:bool</code>
<b>Return Values</b>	-

**extract\_heuristic\_graphs()**

---

<b>Description</b>	Βάσει μεθόδου και συγκεκριμένων ιδιοτήτων επιλέγει την αντίστοιχη σημειογραφία και με την χρήση του Ευρετικού αλγορίθμου παράγει ένα χάρτη διαδικασιών ή ένα Ευρετικό Δίκτυο. Το εμφανίζει στο χρήστη και παράλληλα το αποθηκεύει σε .png αρχείο στο φάκελο που του επιδεικνύει ο χρήστης.
<b>Prototype</b>	<code>extract_heuristic_graphs(method: str, log, filepath: str, _dependency=0.5, _and=0.5, _loop_two=0.5, _act_count=1, _dfg_occurrences=1, _dec="frequency").</code>
<b>Parameters</b>	<code>method:str, log:dataframe, filepath:str, _dependency:float, _and:float, _loop_two:float, _act_count:int, _dfg_occurrences:int, _dec:str</code>
<b>Return Values</b>	-

**discover\_high\_level\_process()**


---

<b>Description</b>	Δημιουργήθηκε για το dataset με το οποίο ασχολείται το συγκεκριμένο κείμενο με σκοπό να ομαδοποιήσει τις δραστηριότητες της διαδικασίας σε επίπεδο εγγράφου και να αποτυπώσει την διαδικασία σε υψηλό επίπεδο με τις σημειογραφίες BPMN και PetriNet.
<b>Prototype</b>	<code>discover_high_level_process(event_log:dataframe).</code>
<b>Parameters</b>	<code>event_log:dataframe,</code> <code>filepath:str</code>
<b>Return Values</b>	<code>event_log:dataframe</code>

**7.2.6 Έλεγχος συμμόρφωσης****diagnostic\_results\_alignment\_method()**


---

<b>Description</b>	Δέχεται σαν είσοδο ένα PetriNet και βάσει παραμέτρων κάνει έλεγχο συμμόρφωσης με την τεχνική "Alignments". Επιπλέον εξάγει χρήσιμα στατιστικά για την διαδικασία και αξιολογεί το μοντέλο εισόδου βάσει των τεσσάρων κριτηρίων αξιολόγησης.
<b>Prototype</b>	<code>diagnostic_results_alignment_method(method:str,</code> <code>log, net, initial_marking, final_marking,</code> <code>enable_separation:bool=False).</code>
<b>Parameters</b>	<code>method:str,</code> <code>log:dataframe,</code> <code>net,</code> <code>initial_marking,</code> <code>final_marking,</code> <code>enable_separation:bool</code>
<b>Return Values</b>	-

**footprints\_compare()**


---

<b>Description</b>	Πρόκειται για μία τεχνική που, ανάλογα τα ορίσματα της, εφαρμόζει έλεγχο συμμόρφωσής με την τεχνοτροπία των footprints στο μοντέλο διαδικασιών που δέχεται σαν είσοδο. Για παράδειγμά σε ένα PetriNet ή ένα Process Tree.
<b>Prototype</b>	<code>footprints_compare(method, event_log,</code> <code>petrinet:list, tree=None).</code>
<b>Parameters</b>	<code>method:str,</code> <code>event_log:dataframe,</code> <code>petrinet:list,</code> <code>process_tree:dict</code>
<b>Return Values</b>	-

---

**diagnostic\_results\_token\_replay\_method()**


---

<b>Description</b>	Δέχεται σαν είσοδο ένα PetriNet και βάσει παραμέτρων κάνει έλεγχο συμμόρφωσης με την τεχνική "Token Replay". Επιπλέον εξάγει χρήσιμα στατιστικά για την διαδικασία και αξιολογεί το μοντέλο εισόδου βάσει των τεσσάρων κριτηρίων αξιολόγησης.
<b>Prototype</b>	<code>diagnostic_results_token_replay_method(method: str, net, initial_marking, final_marking, k_top_variants_log, original_log, enable_separation:bool=False, extra_diagnostics:bool=False).</code>
<b>Parameters</b>	<code>method:str, net, initial_marking, final_marking, k_top_variants_log:dataframe, original_log:dataframe, enable_separation:bool, extra_diagnostics:bool</code>
<b>Return Values</b>	-

---

**extra\_tbr\_diagnostics()**


---

<b>Description</b>	Εκτελεί επιπλέον διαγνωστικούς ελέγχους σχετικά με το αποτέλεσμα που προκύπτει από τον έλεγχο συμμόρφωσης με την τεχνική του "Token Replay". Πληροφορίες που παρέχονται είναι η ανάλυση ρυθμού επεξεργασίας στις μεταβάσεις που εκτελούνται με εσφαλμένο τρόπο σύμφωνα με το μοντέλο διεργασίας (το Petri net), η ανάλυση ρυθμού επεξεργασίας στις δραστηριότητες που δεν περιέχονται στο μοντέλο, η ανάλυση των αιτιών που οδηγούν σε εσφαλμένη εκτέλεση των μεταβάσεων και η ανάλυση των αιτιών που οδηγούν στην εκτέλεση δραστηριοτήτων που δεν περιέχονται στο μοντέλο διεργασίας.
<b>Prototype</b>	<code>extra_tbr_diagnostics(log, net, initial_marking, final_marking, filtered_log, diagnostic_method: list).</code>
<b>Parameters</b>	<code>log:dataframe, net, initial_marking, final_marking, filtered_log:dataframe, diagnostic_method:list</code>
<b>Return Values</b>	-

**diagnostic\_results\_alignment\_method\_change\_classifier()**


---

<b>Description</b>	Κάνει έλεγχο συμμόρφωσης με την τεχνική των ευθυγραμμίσεων μεταξύ ενός δένδρου διαδικασιών και του αρχείου καταγραφής γεγονότων χρησιμοποιώντας custom .
<b>Prototype</b>	diagnostic_results_token_replay_method(event_log).
<b>Parameters</b>	event_log:dataframe
<b>Return Values</b>	-

**7.2.7 Βελτιστοποίηση και ανάλυση****graph\_analytics()**


---

<b>Description</b>	Δέχεται σαν όρισμα ένα γράφο με την μορφή SNA και εκτελεί βασικές πράξεις της θεωρίας γραφημάτων (degree, centrality κα).
<b>Prototype</b>	graph_analytics(pm4py_sna).
<b>Parameters</b>	pm4py_sna:SNA
<b>Return Values</b>	-

**resource\_activity\_matrix()**


---

<b>Description</b>	Δέχεται σαν όρισμα ένα αρχείο καταγραφής δεδομένων και εκτυπώνει και αποθηκεύει σε .csv αρχείο ένα πίνακα δραστηριοτήτων.
<b>Prototype</b>	resource_activity_matrix(event_log, extract_to_csv:bool = False).
<b>Parameters</b>	event_log:dataframe, extract_to_csv:bool
<b>Return Values</b>	resource_activity_matrix:dataframe

**work\_handover\_matrix()**


---

<b>Description</b>	Δέχεται σαν όρισμα ένα αρχείο καταγραφής δεδομένων και εκτυπώνει και αποθηκεύει σε .csv αρχείο έναν πίνακα "Work Handover".
<b>Prototype</b>	work_handover_matrix(event_log, extract_to_csv:bool=False).
<b>Parameters</b>	event_log:dataframe, extract_to_csv:bool
<b>Return Values</b>	handover_matrix:dataframe

**discovering\_roles()**

---

<b>Description</b>	Ανακαλύπτει τους ρόλους που έχουν τα resources σε ένα αρχείο καταγραφής γεγονότων .
<b>Prototype</b>	discovering_roles(event_log).
<b>Parameters</b>	event_log:dataframe
<b>Return Values</b>	roles:dataframe

**decision\_tree\_mining()**

---

<b>Description</b>	Κατασκευάζει το Decision Tree και το εκτυπώνει σε μορφή .pdf ανάλογα τις response και predictor μεταβλητές που δέχεται σαν όρισμα .
<b>Prototype</b>	decision_tree_mining(dataframe_event_log, response_value:str, independent_value:list).
<b>Parameters</b>	dataframe_event_log:dataframe, response_value:str, independent_value:list
<b>Return Values</b>	-





# Εφαρμογή σε διαδικασίες χρηματοδότησης της Κοινής Αγροτικής Πολιτικής

## 8.1 Παρουσίαση Δεδομένων

Αρχικά γίνεται αναφορά στο σύνολο των περιπτώσεων και στο σύνολο των γεγονότων που περιέχεται μέσα στο αρχείο καταγραφής γεγονότων.

Table 8.1: Δεδομένα αρχείου καταγραφής γεγονότων.

Total Traces	Total Events
43809	2514266

Το επόμενο γράφημα παρουσιάζεται η κατανομή των 43809 αιτήσεις έχουν ληφθεί τις χρονιές 2015, 2016 και 2017 και παρατηρείται ότι ο αριθμός των αιτήσεων ανά χρονιά είναι σχεδόν σταθερός.

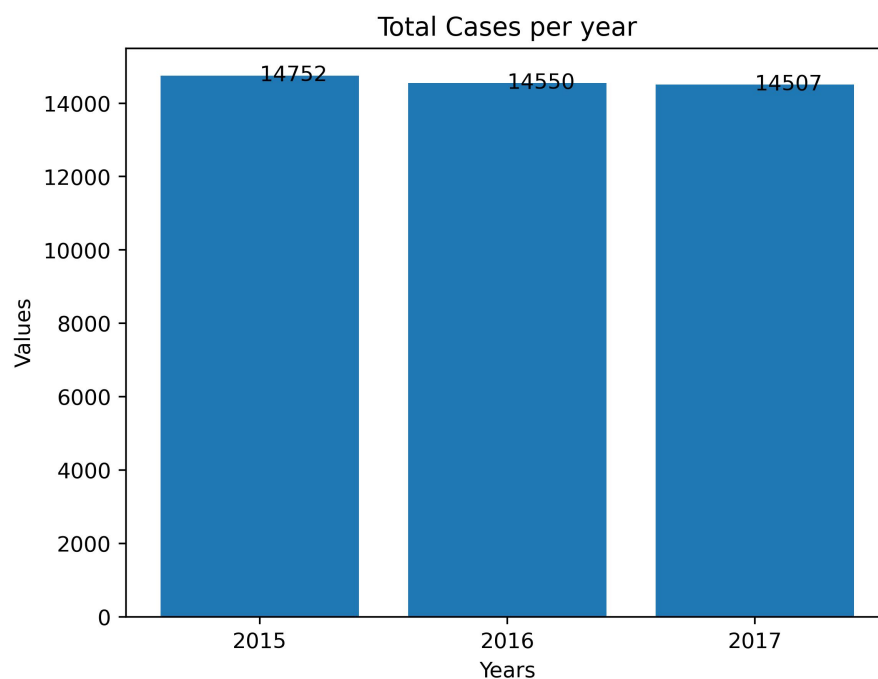


Figure 8.1: Κατανομή των αιτήσεων ανά έτος.

Στην συνέχεια γίνεται εξαγωγή του μήνα κατά τον οποίο κατατίθεται μία αίτηση μέσα στην χρονιά. Έτσι παρατηρείται ότι οι περισσότερες περιπτώσεις ξεκινούν τους μήνες του Απριλίου και του Μαΐου, με τον Μάιο να έχει τις περισσότερες, υπάρχον βέβαια και στην διάρκεια των υπολοίπων μηνών κάποιες αιτήσεις, άλλα το δείγμα είναι αρκετά μικρό ώστε να βγει κάποιο συμπέρασμα.

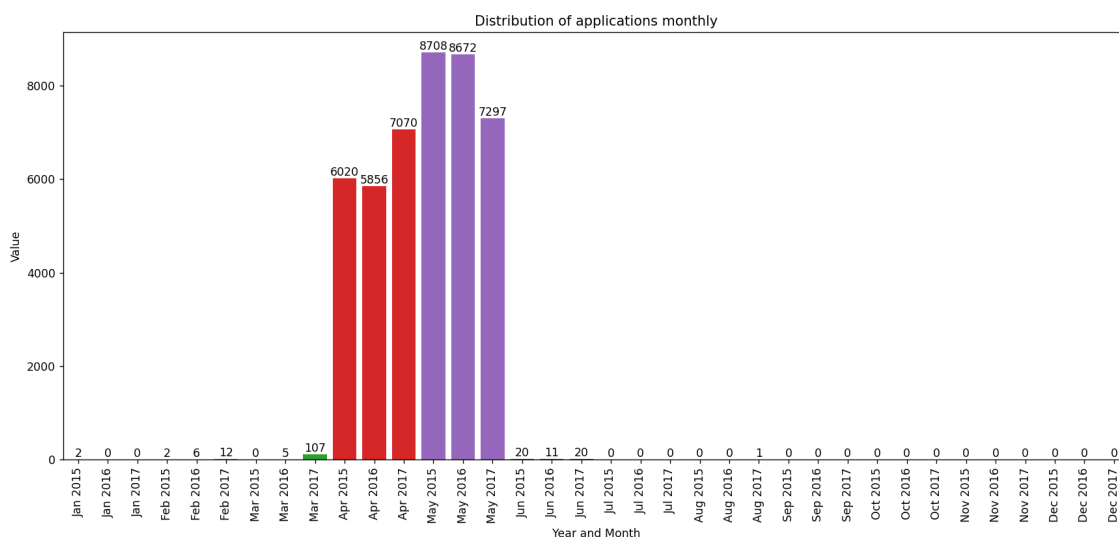


Figure 8.2: Κατανομή των αιτήσεων ανά μήνα και έτος.

Σε επίπεδο γεγονότων και δραστηριοτήτων στο σύνολο των 43809 οι δραστηριότητες που αρχίζουν την διαδικασία είναι τέσσερις και οι δραστηριότητες οι οποίες τελειώνουν μια διαδικασία είναι είκοσι μία. Σύμφωνα με το documentation, η αρχική δραστηριότητα είναι *mail income* του εγγράφου *Payment Application* με 38623 περιπτώσεις να έχουν αρχίσει με την συγκεκριμένη και η τελική δραστηριότητα είναι *finish payment* του ίδιου εγγράφου με 34830 περιπτώσεις να ολοκληρώνονται με την συγκεκριμένη. Οι υπόλοιπες δραστηριότητες που έχουν ανακαλυφθεί να αρχίζουν με διαφορετικό τρόπο οφείλονται σε χειριστικά λάθη και ελλιπή καταγραφή των γεγονότων από το πληροφοριακό σύστημα.

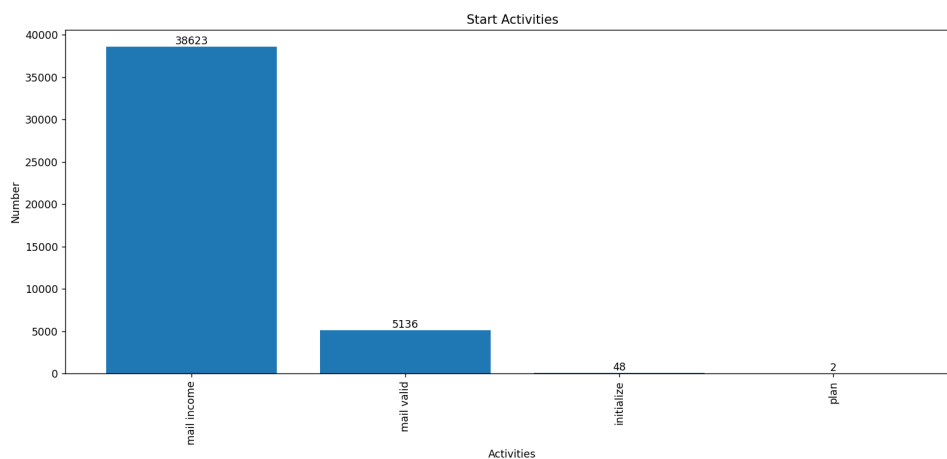


Figure 8.3: Δραστηριότητες Έναρξης

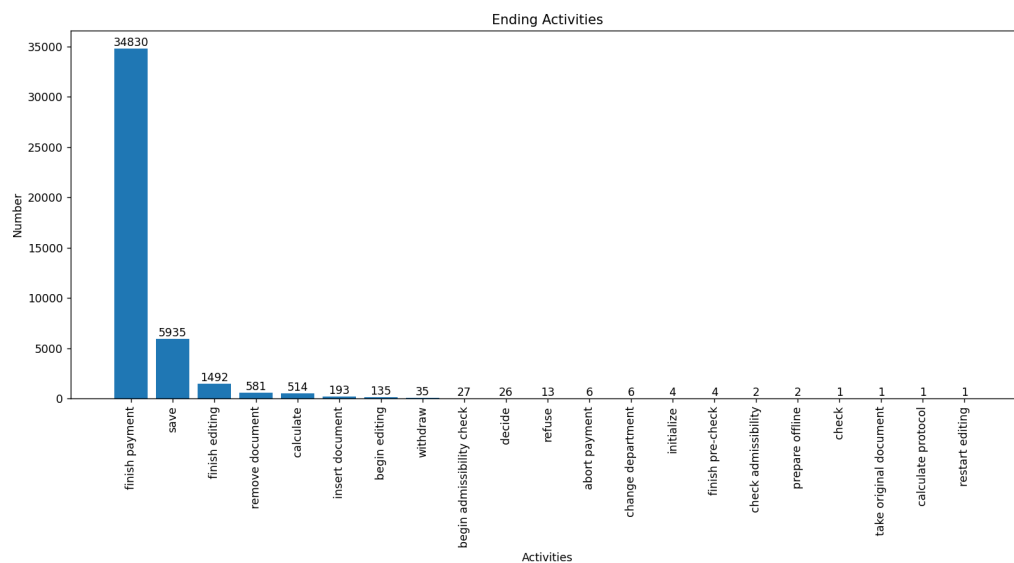


Figure 8.4: Δραστηριότητες Ολοκλήρωσης

Από το συνολικό αρχείο καταγραφής μπορεί να εξαχθεί το πότε άρχισε και τελείωσε μία διαδικασία, δηλαδή την χρονική στιγμή με βάση το έτος και κάποια από τις δραστηριότητες έναρξης που κατατίθεται μία αίτηση και το πότε τελειώνει με βάση το έτος και την τελική δραστηριότητα *finish payment*.

Table 8.2: Έτη έναρξης και ολοκλήρωσης διαδικασιών.

Έτος Έναρξης	Έτος Ολοκλήρωσης	Αριθμός Δραστηριοτήτων
2015	2015	0
2015	2016	14746
2015	2017	3
2015	2018	0
2016	2016	0
2016	2017	14545
2016	2018	0
2017	2017	0
2017	2018	14427

Γίνεται ανάκτηση της λίστας με το σύνολο της διάρκειας για κάθε περίπτωση ξεχωριστά (η αρχική ανάκτηση της πληροφορίας γίνεται σε δευτερόλεπτα και μετατρέπεται σε μέρες) και υπολογίζεται η μεγαλύτερη, η μικρότερη και ο μέσος όρος των περιπτώσεων στο συνολικό αρχείο καταγραφής δεδομένων και ανά χρόνια υποβολής αίτησης. Από αυτό το σημείο και μετά προκύπτουν αρκετά συμπεράσματα σχετικά με την χρονική διάρκεια ολοκλήρωσης της διαδικασίας επεξεργασίας των αιτήσεων στα διαφορετικά έτη που έχουμε σαν δεδομένα.

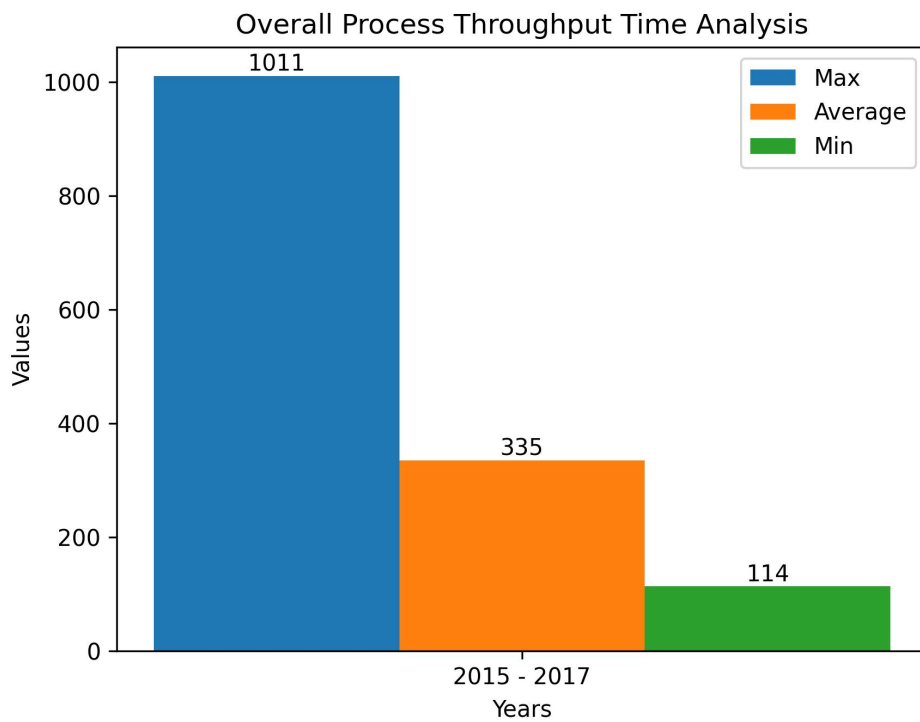


Figure 8.5: Συνολική Διάρκεια περιπτώσεων

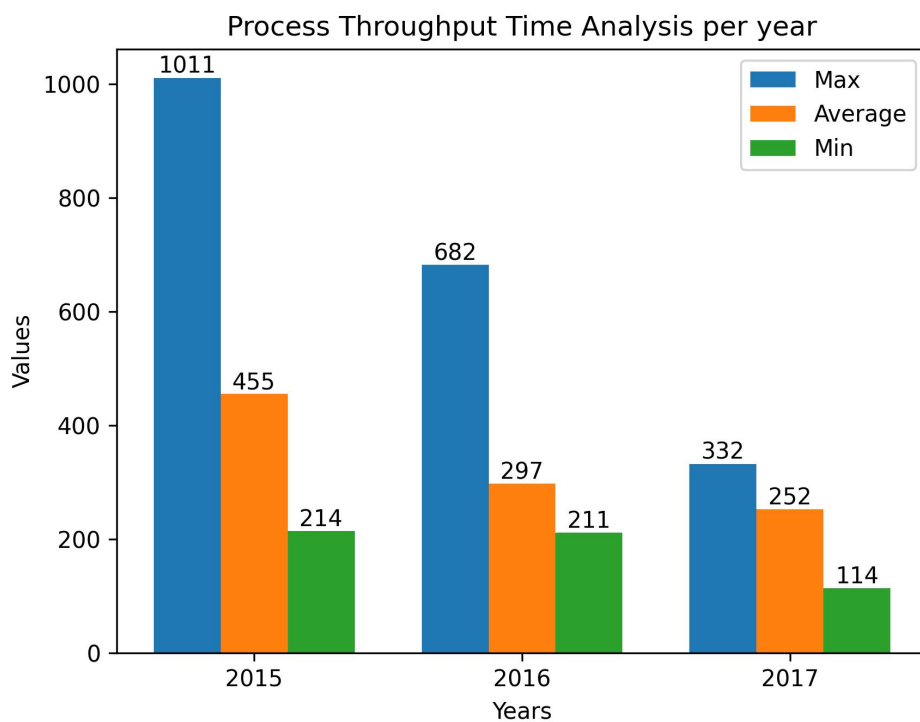


Figure 8.6: Διάρκεια περιπτώσεων ανά χρονιά

Σε γενικές γραμμές από την περιγραφή<sup>1</sup> της διαδικασίας και τα πρώτα στοιχεία, που έχουν αντληθεί από το αρχείο καταγραφής δεδομένων, προκύπτει ότι το πρόγραμμα πληρωμών με την συγκεκριμένη μορφή τέθηκε σε ισχύ το 2015. Την πρώτη χρονιά που τέθηκε σε εφαρμογή ελήφθησαν από το σύστημα συνολικά 14752 αιτήσεις. Τις επόμενες χρονιές (2016, 2017) ο αριθμός αυτός εμφανίστηκε ελάχιστα μειωμένος, αλλά όχι σε σημαντικό βαθμό. Μετά από εξέταση των αγροτών, βάσει του αναγνωριστικού κωδικού του αιτούντος αγρότη, διαπιστώθηκε ότι δεν διαφοροποιούνται παρά ελάχιστα μέσα στις χρονιές και πρόκειται ουσιαστικά για τους ίδιους αγρότες, το οποίο επιβεβαιώνεται από τον πίνακα 8.3 που δείχνει ακριβώς τον αριθμό των αιτούντων που επαναλαμβάνουν την αίτηση τα επόμενα έτη. Διαπιστώνεται επίσης, ότι οι αιτήσεις έγιναν το ίδιο χρονικό διάστημα μέσα στην χρονιά, δηλαδή τους μήνες Απρίλιο και Μάιο με ελάχιστες εξαιρέσεις στις 43809 περιπτώσεις που έχουν καταγραφεί στο αρχείο.

Table 8.3: Αριθμός αιτούντων ανά τα χρόνια αιτήσεων.

Applicants during 2015 and 2016	13908
Applicants during 2015 and 2017	13361
Applicants during 2016 and 2017	13906

Επίσης, βάσει των χαρακτηριστικών σε επίπεδο διαδικασίας υπολογίζεται η κατανομή εργασίας που έχει κάθε τμήμα, διαπιστώνοντας από την εικόνα 8.7 ότι τρία από τα τμήματα έχουν σχεδόν το ίδιο φόρτο εργασίας και ένα με λιγότερες αιτήσεις από ότι ο μέσος όρος των άλλων.

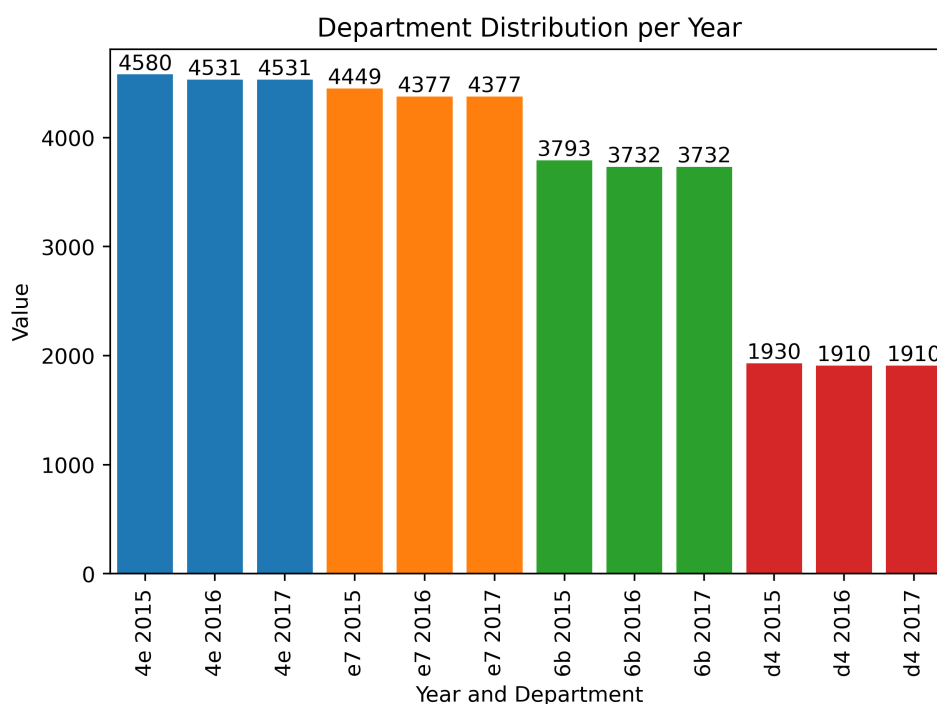


Figure 8.7: Κατανομή εργασιών ανά τμήμα και χρονιά.

<sup>1</sup><https://www.win.tue.nl/bpi/2018/challenge.html>

Table 8.4: Κατανομή ποινών ανά τα έτη 2015, 2016, 2017

Κατηγορία Ποινής	2015	2016	2017
AJLP	22	20	19
BGKV	0	0	15
AUVP	367	251	211
BGP	0	0	372
C16	1	1	3
BGK	0	119	174
AVUVP	0	2	0
CC	528	411	404
AVJLP	1	2	0
C9	17	5	4
C4	192	156	0
AVGP	1	1	0
ABP	248	245	211
B6	123	0	0
B4	191	0	0
B5	0	51	105
AVBP	1	1	0
B2	7967	2058	1502
B3	805	0	0
AGP	244	247	211
B16	111	0	0
GP1	795	614	916
B5F	0	0	9
V5	0	1	0
JLP6	0	58	66
JLP7	0	0	1
JLP5	0	11	14
JLP2	32	8	5
JLP3	2	0	10
JLP1	99	45	25

Η εξόρυξη των δεδομένων [8.4] που αφορούν τις ποινές ανά έτος μας επιβεβαιώνει ότι πρόκειται για μια διαδικασία επαναλαμβανόμενη από τα ίδια άτομα κάθε χρόνο. Η επανάληψη δείχνει ότι ο συνολικός αριθμός των ποινών μειώνεται σε γενικές γραμμές και σε συνδυασμό με τους μέσους όρους ολοκλήρωσης [8.6] της διαδικασίας, προκύπτει το συμπέρασμα ότι οι χρήστες έ"εμπαιδευτήκαν" στον τρόπο υποβολής της αίτησης με την παρόδο των ετών. Επίσης, οι αλλαγές που έγιναν από την πλευρά του οργανισμού Κοινής Αγροτικής Ανάπτυξης, δηλαδή η αντικατάσταση των διαδικασιών που περιγράφουν οι τύποι εγγράφων *Parcel Document* (πριν το 2016), *Department control parcels* (πριν το 2017), και *Geo Parcel Document* (μετά το 2017) και σε συνδυασμό με την καλύτερη εικόνα των δεδομένων, δείχνουν βελτίωση της διαδικασίας στην διάρκεια αυτών των τριών ετών.

Table 8.5: Αριθμός δραστηριοτήτων εκκίνησης - τερματισμού ανά έτος.

Year	Δραστηριότητες Εκκίνησης	Δραστηριότητες Τερματισμού
2015	3	13
2016	4	17
2017	3	10

Τέλος, από τα αποτελέσματα που προέκυψαν και συγκεκριμένα τον μέσο όρο ολοκ-

λήρωσης μια διαδικασίας[8.6], την άνιση κατανομή δραστηριοτήτων τερματισμού ανά έτος[8.5] και το συμπέρασμά ότι σε αρκετές περιπτώσεις απουσιάζουν συγκεκριμένες δραστηριότητες που καθορίζουν το τέλος μιας αίτησης είναι αδύνατο να προσδιοριστεί ο χρόνος ολοκλήρωσης όλων των περιπτώσεων. Επίσης, τα δεδομένα που έχουν καταγραφεί στο αρχείο είναι διαθέσιμα μέχρι τις 19 Ιανουαρίου 2018, γίνεται αντιληπτό ότι για κάθε έτος έχουμε διαθέσιμο διαφορετικό χρονικό διάστημα. Έτσι, για τις αιτήσεις που υποβλήθηκαν την χρονική περίοδο Απριλίου, Μαΐου του 2017, υπάρχει ιστορικότητα περίπου 9 μηνών. Από την άλλη πλευρά, για αιτήσεις που υποβλήθηκαν το 2015, έχουμε ιστορικό γεγονότων για περισσότερο από 2 χρόνια και δεν είναι τυχαίο ότι υπάρχουν περιπτώσεις τον Ιανουάριο του 2018 που εκκρεμούν από το 2015. Όλες οι παραπάνω παρατηρήσεις θα επηρεάσουν αρνητικά τις υπόλοιπες εργασίες εξόρυξης διαδικασιών και συγκεκριμένα τους αλγόριθμους εξόρυξης, όπως ο αλγόριθμος-*a*, και συμμόρφωσης, όπως τα footprints, οι οποίοι δεν μπορούν να διαχειριστούν τον θόρυβο και τα ατελή δεδομένα.

## 8.2 Ανακάλυψη Διαδικασίας

Στο επόμενο βήμα θα γίνει η ανακάλυψη της διαδικασίας. Από την ανάλυση του αρχείου έχει προκύψει το συμπέρασμα ότι το αρχείο καταγραφής γεγονότων έχει αρκετό θόρυβο και αρκετές από τις περιπτώσεις δεν έχουν ολοκληρωθεί, όπως υποδεικνύει η καταμέτρηση των διαφορετικών περιπτώσεων εκτέλεσης της διαδικασίας (variants) και το πλήθος των δραστηριοτήτων τερματισμού. Συγκεκριμένα προκύπτει ότι υπάρχουν 28457 variants από τα οποία τα 26602 έχουν εκτελεστεί μόνο μία φορά, δηλαδή το 93.48%. Για αυτό τον λόγο και για να γίνει η διαδικασία πιο κατανοητή στον αναγνώστη η εργασία του Process Discovery και του Conformance Checking θα γίνει με τις 10 πιο κοινές περιπτώσεις οι οποίες αποτελούν το 18.9%(5377) του συνολικού αρχείου καταγραφής γεγονότων και περιέχουν 237492 events.

Πριν όμως την ανακάλυψη της διαδικασίας σε επίπεδο δραστηριοτήτων και λόγω της ανάγκης να γίνει κατανοητή στον αναγνώστη που δεν είναι εξοικειωμένος με την ανάγνωση και κατανόηση των διαγραμμάτων που αναπαριστούν τα μοντέλα διαδικασιών, προηγήθηκε μια προ-επεξεργασία των δεδομένων. Αρχικά, έγινε συγχώνευση της διαδικασίας σε επίπεδο τύπου-εγγράφου ανάλογα με την έναρξη και ολοκλήρωσης κάποιας από τις δραστηριότητες που τα απαρτίζουν και στην εξόρυξη μοντέλων διαδικασίας με την χρήση του αλγόριθμου "inductive miner" με αποτέλεσμα να εξαχθούν μοντέλα διαδικασιών που περιγράφουν την διαδικασία σε υψηλό επίπεδο. Πρόκειται για ένα μοντέλο BPMN(εικόνα 8.8) και έναν γράφο *Directly Follow Graph* (εικόνα 8.9) με σχόλια αναφορικά με το performance των διαδικασιών. Τα δύο αυτά μοντέλα παρουσιάζουν την ροή εκτέλεσης της διαδικασίας, πώς αρχίζει να ολοκληρώνεται και ποια έγγραφα περιέχουν δραστηριότητες οι οποίες εκτελούνται είτε συνεχόμενα, είτε διακόπτονται για να συνεχιστούν αργότερα, αφού χρονικά προηγούνται δραστηριότητες άλλων εγγράφων. Από την ανάλυση του χάρτη διαδικασίας (εικόνα 8.9) ο οποίος δημιουργήθηκε παρατηρούμε ότι τα έγγραφα πλέον απεικονίζουν τις δραστηριότητες, οι ακμές την μεταξύ τους σχέση και οι αριθμοί την διάρκεια μετάβασής από την μια δραστηριότητα στην άλλη. Αναλύοντας σε βάθος είναι εύκολο να διακρίνουμε ότι υπάρχουν πολλοί βρόχοι στη διαδικασία. Η πιο σημαντική παρατήρηση αναφορικά με αυτούς τους βρόχους είναι ότι συνήθως ολόκληρη η διαδικασία ξεκινά και τελειώνει με τη δραστηριότητα "Payment Application". Μια άλλη παρατήρηση είναι η επανάληψη των δραστηριοτήτων. Αυτό οφείλεται στο γεγονός ότι το συγκεκριμένο μοντέλο



διαδικασιών βασίζεται μόνο στους τύπους εγγράφων και όχι στις δραστηριότητες που το αποτελούν. Η ανάλυση των δραστηριοτήτων παρακάτω, δείχνει ότι υπάρχουν διαφορές στον αριθμό των περιπτώσεων που έχουν πραγματοποιήσει τη δραστηριότητα. Επίσης παρατηρούμε ότι πολλά βέλη τερματίζουν από δραστηριότητες πλην του "Payment Application" και αυτό έρχεται να επιβεβαιώσει την εύρεση του μεγάλου αριθμού δραστηριοτήτων τερματισμού η οποία οφείλεται κυρίως σε ελλιπή καταγραφή γεγονότων.

Στην συνέχεια οι υπόλοιπες εργασίες ανακάλυψης και ανάλυσης θα γίνουν σε επίπεδο δραστηριοτήτων με την χρήση των αλγορίθμων εξόρυξης διαδικασιών, ώστε να εξαχθούν ορθότερα συμπεράσματα για την ροή εκτέλεσής τους. Εκτελώντας τον αλγόριθμο-α δεν προέκυψε κάποιο μοντέλο για περαιτέρω ανάλυση. Στο διάγραμμα PetriNet που ανακαλύπτει εμφανίστηκαν μόνο οι δραστηριότητες σε μορφή μεταβάσεων ενός διαγράμματος PetriNet, χωρίς κάποια ουσιαστική σύνδεση μεταξύ τους. Αυτό οφείλεται κυρίως στο ότι ούτε ο αλγόριθμός ούτε και το API μπορούν να διαχειριστούν δεδομένα που είναι ατελή, περιέχουν θόρυβο και διαδικασίες των οποίων οι σχέσεις των δραστηριοτήτων περιέχουν self και short loops.

Τα επόμενα μοντέλα παρουσιάζουν την διαδικασία με διαφορετικές σημειογραφίες, εκμεταλλεύονται τις ιδιότητες κάθε αλγόριθμου και τεχνικής με σκοπό να παρουσιάσουν την διαδικασία με τον βέλτιστο τρόπο και τις δυνατότητες της *pm4py* ως προς την εφαρμογή των τεχνικών που υποστηρίζει. Εφαρμόζονται οι αλγόριθμοι ανακάλυψης διαδικασιών 'inductive miner', 'heuristic miner' και η τεχνική ανακάλυψης. Παρουσιάζονται τα μοντέλα αναλυτικά,

- *Heuristic miner-PetriNet-Top 10 variants, Εικόνα 8.11*, Παράγεται ένα PetriNet το οποίο εξάγεται με τον Ευρετικό αλγόριθμο ο οποίος εφαρμόζεται στις 10 δημοφιλέστερες περιπτώσεις της διαδικασίας που περιέχει το αρχείο καταγραφής γεγονότων. Στην παρούσα εκτέλεση δεν έχει οριστεί κάποιο κατώφλι εμφάνισης των δραστηριοτήτων βάσει της συχνότητας με την οποία εκτελούνται (dependency) Με την εκτέλεση συγκεκριμένων διαγνωστικών της *pm4py* (WOFLAN<sup>2</sup>. Το PetriNet που αποτυπώνει την διαδικασία αξιολογείται από την μέθοδο ανάλυσης WOFLAN ως εξής,

- Input is ok.
- Petri Net is a workflow net.
- The following places are not covered by an s-component: `intplace_begin editing`, `intplace_begin payment`, `intplace_decide`, `intplace_finish editing`, `intplace_initialize`, `intplace_insert document`, `intplace_performed pre_abort payment`, `pre_begin editing`, `pre_begin payment`, `pre_calculate`, `pre_decide`, `pre_finish editing`, `pre_finish payment`, `pre_initialize`, `pre_performed`, `sink0`, `source0`, `splace_in_calculate_calculate_0`, `splace_in_finish editing_calculate_0`, `splace_in_finish editing_finish editing_0`, `splace_in_insert document_insert document_0`, `splace_in_mail valid_mail income_0`.
- There are dead tasks.
- Number of Uncovered Places : 23.

<sup>2</sup>Το WOFLAN είναι μια δημοφιλής προσέγγιση για τον έλεγχο ευρωστίας στα δίχτυα ροής εργασιών, η οποία είναι σε θέση να παρέχει ουσιαστικά στατιστικά στοιχεία στον τελικό χρήστη

Figure 8.8: Process flow with high-level work flow steps - BPMN Process model

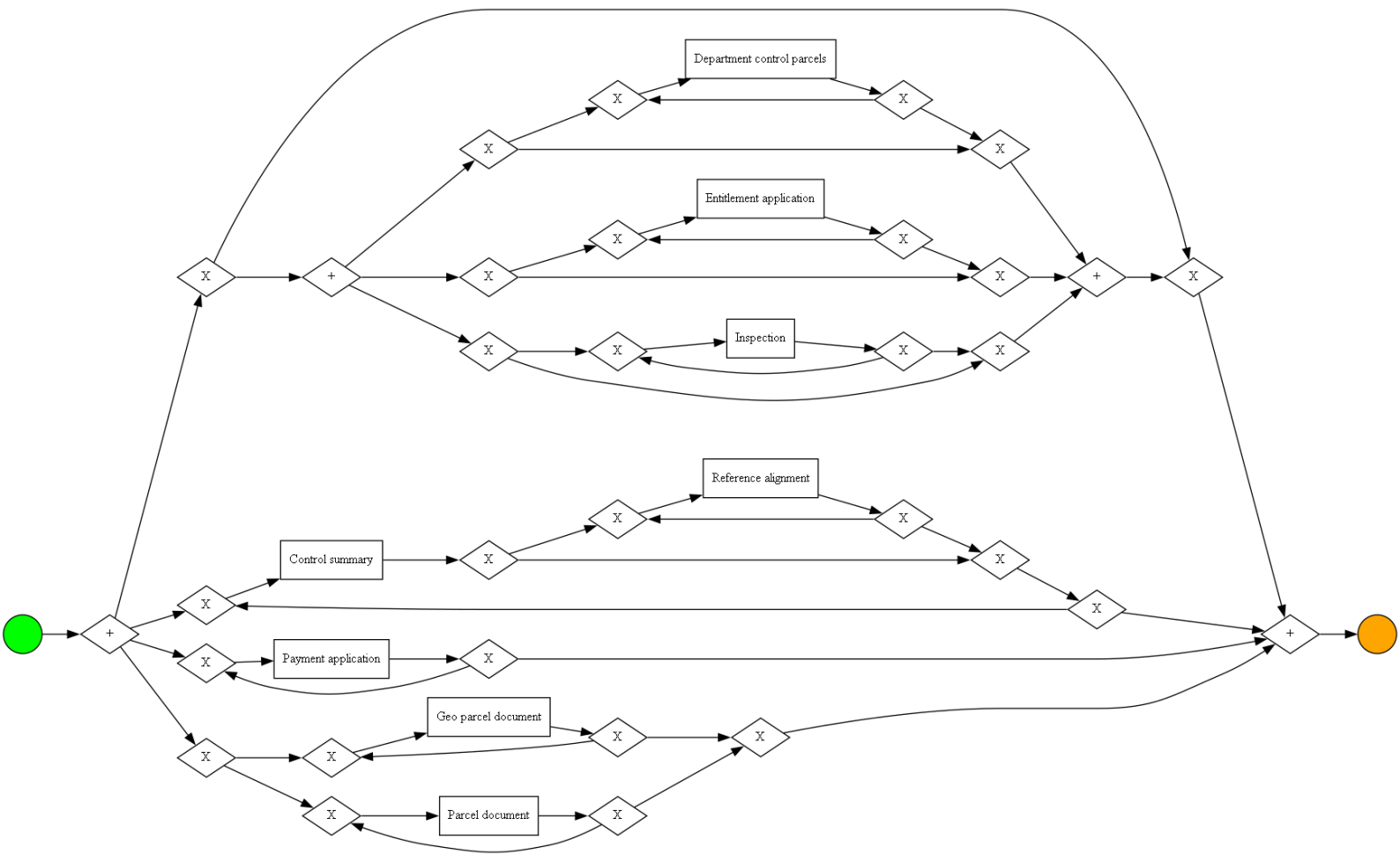
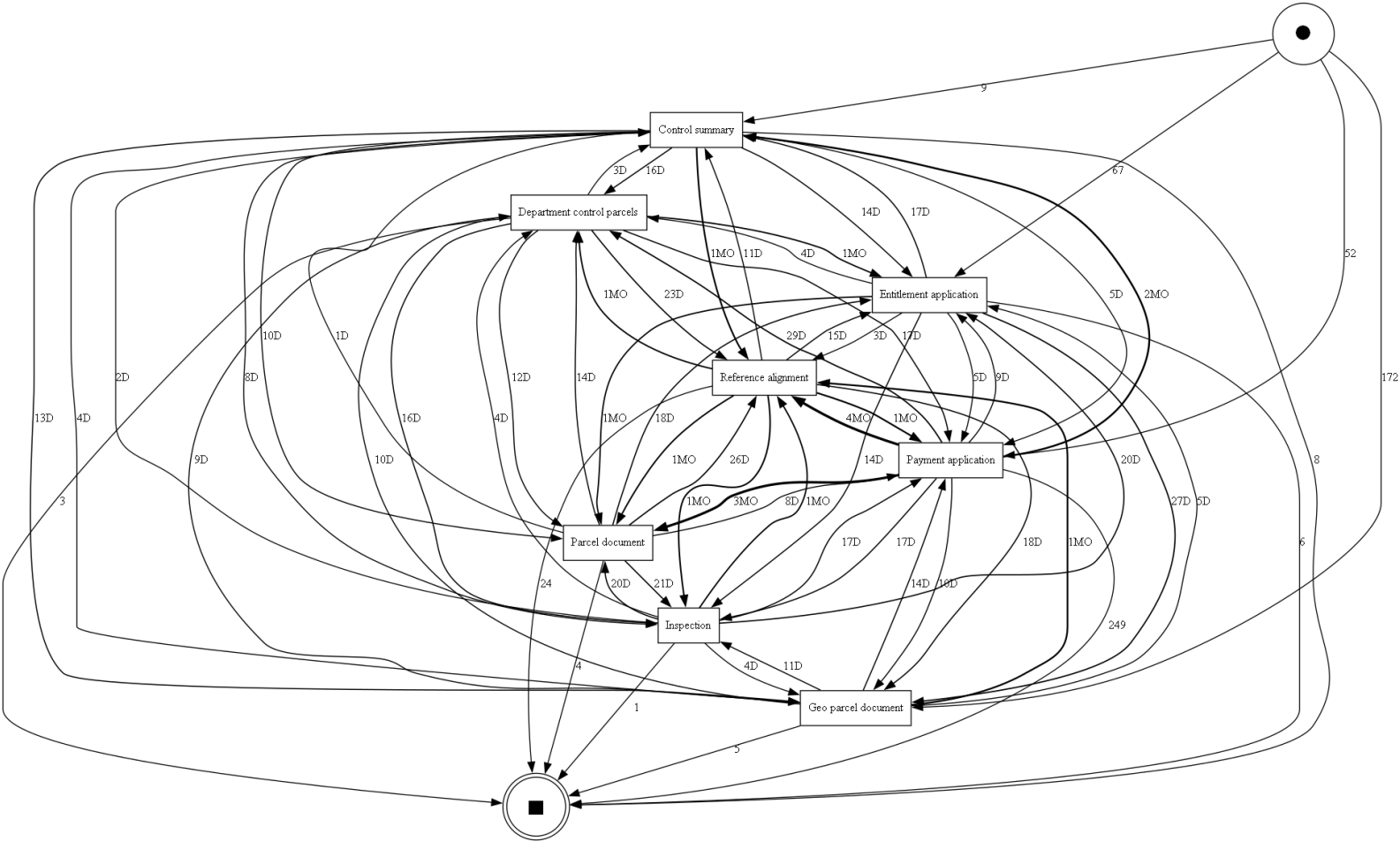


Figure 8.9: Process flow with high-level work flow steps - Directly Follow Graph



- *Heuristic miner - Heuristic Graph - Dependency 0.80, Εικόνα 8.12.* Παράγεται ένα Ευρετικό Δίκτυο το οποίο αφορά το σύνολο του αρχείου καταγραφής γεγονότων. Η τιμή που εκφράζει την εξάρτηση μεταξύ των δραστηριοτήτων έχει τεθεί αρκετά χαμηλά, μόλις 0.80, με αποτέλεσμα να αποκλείεται μεγάλος αριθμός περιπτώσεων. Ακόμα και με αυτόν τον τρόπο όμως το μοντέλο διαδικασιών που έχει ανακαλυφθεί είναι αρκετά δυσνόητο προς κάποιον χρήστη ο οποίος δεν έχει κατάλληλες γνώσεις, γεγονός που επιβεβαιώνει άλλη μια φορά την πολυπλοκότητα της διαδικασίας.
- *Heuristic miner - Heuristic Graph - Top 10 variants, Εικόνες 8.10, 8.13.* Παράγεται ένα Ευρετικό Δίκτυο το οποίο αφορά τις 10 καλύτερες περιπτώσεις του αρχείου καταγραφής γεγονότων. Περιλαμβάνει όλες τις περιπτώσεις του δείγματος του συνόλου του αρχείου καταγραφής, με αποτέλεσμα ένα αρκετά κατανοητό μοντέλο διαδικασιών. Αυτό συμβαίνει γιατί όλες οι περιπτώσεις εκτελούν την συγκεκριμένη διαδικασία αφού αρχίζουν συγκεκριμένη δραστηριότητα έναρξης και ολοκληρώνονται σε συγκεκριμένη δραστηριότητα τερματισμού. Το διάγραμμα είναι εμπλουτισμένο με το performance της διαδικασίας, δηλαδή το χρόνο που χρειάστηκε κάποια δραστηριότητα να ολοκληρωθεί. Έρχεται να επιβεβαιώσει το PetriNet που είχε παραχθεί με τα αντίστοιχα δεδομένα ότι είναι πράγματι ένα WorkFlowNet. Για λόγους επίδειξης δυνατοτήτων της βιβλιοθήκης *pm4py* έχει προστεθεί επίσης ένα Heuristic Net με dependency 0.80, όπου το αποτέλεσμα είναι ένα δίκτυο με λιγότερες μεταβάσεις.
- *Inductive Miner - PetriNet - Top 10 variants - Threshold Noise reduce 40%, Εικόνα 8.14.* Με την χρήση του αλγόριθμου Inductive Miner εκμεταλλευόμαστε το γεγονός ότι μπορεί να διαχειριστεί τον θόρυβο που παράγεται από ένα PetriNet το οποίο αφορά τις 10 καλύτερες περιπτώσεις του αρχείου καταγραφής γεγονότων. Το συγκεκριμένο PetriNet δημιουργήθηκε περισσότερο για λόγους επίδειξης των δυνατοτήτων του αλγόριθμου μειώνοντας τον θόρυβο και αφαιρώντας τις αντίστοιχες περιπτώσεις χρήσεων. Το PetriNet που αποτυπώνει την διαδικασία αξιολογείται από την μέθοδο ανάλυσης WOFLAN ως εξής,
  - Input is ok.
  - Petri Net is a workflow net.
  - Every place is covered by s-components.
  - Number of Uncovered Places : 0.
  - There are no dead tasks.
  - All tasks are live.
- *Inductive Miner - BPMN Model - Top 10 Variants, Εικόνα 8.17.* Με την χρήση του αλγόριθμου "Inductive Miner" παράγεται το αντίστοιχο BPMN μοντέλο. Το BPMN που ανακαλύπτεται είναι αρκετά ευανάγνωστο, οι δραστηριότητες που αποτελούν την ροή της διαδικασίας είναι καθαρές και δείχνει στον αναγνώστη τι ακριβώς συμβαίνει κατά την εκτέλεση της.
- *Inductive Miner - PetriNet - Total Event Log, Εικόνα 8.16.* Με την χρήση του αλγόριθμου Inductive Miner, παράγεται μοντέλο διαδικασιών PetriNet το παρουσιάζει την διαδικασία όπως ακριβώς την καταγράφει το αρχείο χωρίς να

έχει αφαιρέσει ούτε μία περίπτωση από το σύνολο το 43809. Για αυτό το λόγο παράγεται ένα τόσο μεγάλο και πολύπλοκο μοντέλο που αποτυπώνει την συνολική συμπεριφορά που περιέχει το αρχείο καταγραφής γεγονότων. Το PetriNet που αποτυπώνει την διαδικασία αξιολογείται από την *pm4py* ως εξής,

- Input is ok.
- Petri Net is a workflow net.
- Every place is covered by s-components.
- There are no dead tasks.
- All tasks are live.

Η αξιολόγηση με την μέθοδο WOFLAN είναι και η μόνη η οποία θα γίνει στο συγκεκριμένο μοντέλο εξαιτίας της έλλειψης υπολογιστικής ισχύς που χρειάζονται οι υπόλοιπες τεχνικές ελέγχου συμφόρησης.

- *Inductive Miner - Process Tree - Top 10 Variants, Εικόνα 8.18* Παρουσιάζεται το δέντρο διαδικασιών το οποίο είναι αποτέλεσμα υπολογισμών του "Inductive Miner". Ο συγκεκριμένος αλγόριθμος έχει σαν αρχή την παραγωγή του συγκεκριμένου δένδρου και στην συνέχεια η κατασκευή οποιοδήποτε άλλου μοντέλου που παρουσιάστηκε παραπάνω και έχει παραχθεί από τον συγκεκριμένο αλγόριθμο. Είναι από κατασκευής χωρίς αδιέξοδα και τυφλά σημεία και αποτελεί την ιδανική λύση για την ανακάλυψη πολύπλοκων διαδικασιών.
- *Inductive Miner - Reachability Graph - Top 10 Variants, Εικόνα 8.18* ο συγκεκριμένος γράφος είναι ένα σύστημα μεταβάσεων που είναι αποτέλεσμα ανάλυσης του PetriNet και βασική ιδιότητα του είναι ένα σημείο εκκίνησης και έναν μη καθορισμένο αριθμό τελικών σημείων τερματισμού. Διακρίνονται οι διαφορές και οι σημαντικές καταστάσεις που είναι προσβάσιμες από την αρχική κατάσταση. Κάθε κόμβος στο γράφημα αντιπροσωπεύει μια κατάσταση, που είναι μια κατάσταση του συστήματος η οποία αναπαρίσταται από το Petri net και οι ακμές μεταξύ των κόμβων αντιπροσωπεύουν μεταβάσεις που μπορούν να ενεργοποιηθούν για να γίνει η μετακίνηση από μια κατάσταση σε μια άλλη. Επιτρέπει την ανάλυση της συμπεριφοράς του συστήματος, εντοπίζει τις δυνατές ακολουθίες μεταβάσεων και βοηθάει στην κατανόηση της συνολικής δομής των εφικτών καταστάσεων.
- *Directly Follows Graph - Top 10 Variants - Frequency Annotated, Εικόνα 8.20* και *Directly Follows Graph - Top 10 Variants - Performance Annotated, Εικόνα 8.21* Πρόκειται για δύο γράφους, οι οποίοι περιγράφουν την ίδια διαδικασία, έχουν ίδια ροή εκτέλεσης δραστηριοτήτων και αλλάζουν μόνο τον σχολιασμό πάνω στον γράφο. Ο ένας είναι εμπλουτισμένος με δεδομένα που αφορούν την απόδοση και ο άλλος με στοιχεία που αφορούν την συχνότητα. Μπορούν και παρουσιάζουν καθαρά την ροή εκτέλεσης της διαδικασίας και τις σχέσεις μεταξύ των δραστηριοτήτων.

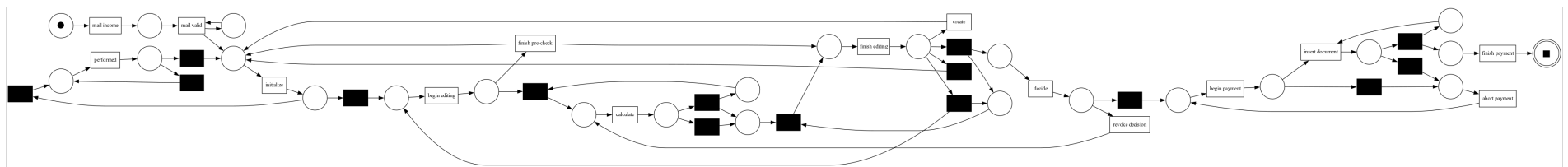
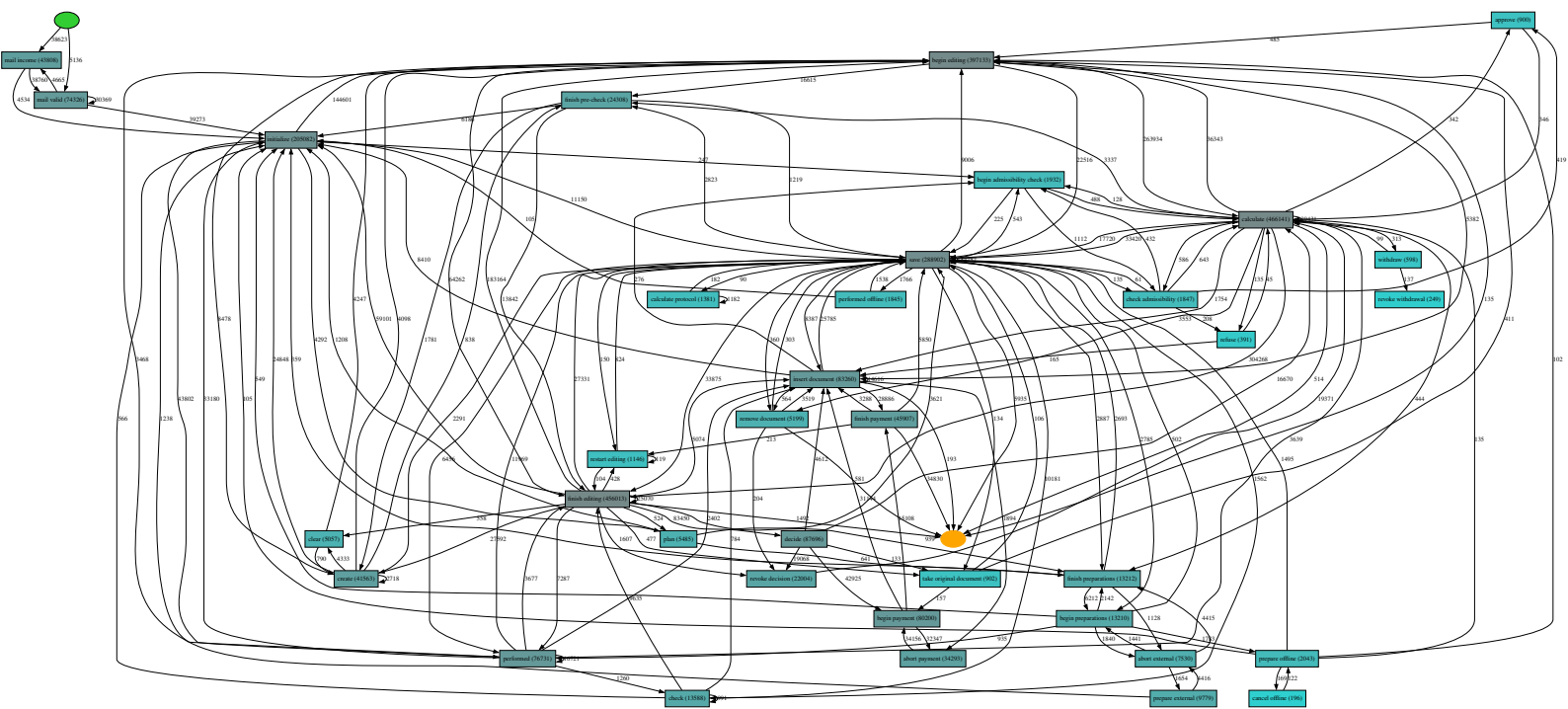


Figure 8.11: PetriNet - Heuristic Miner

Figure 8.12: Heuristic Net - Process Flow Discovery with dependency 0.80



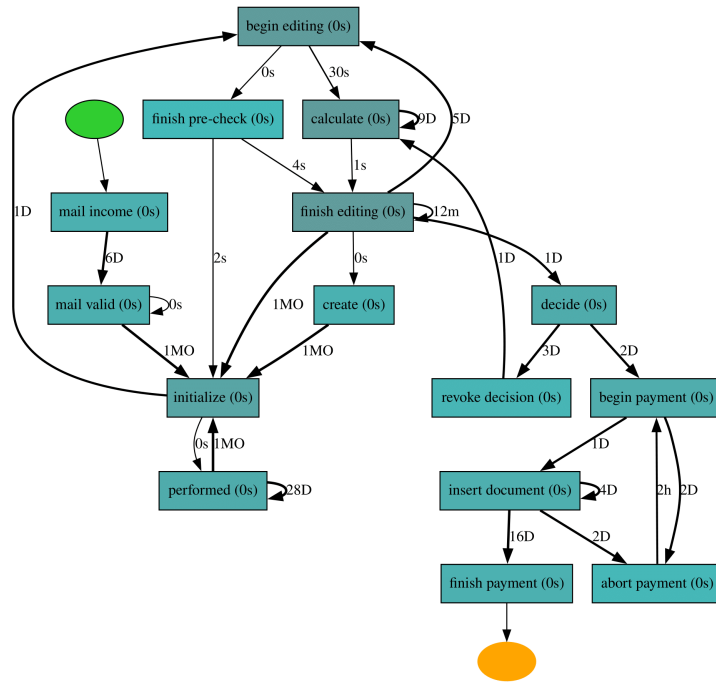


Figure 8.10: Heuristic Net - Top 10 variants

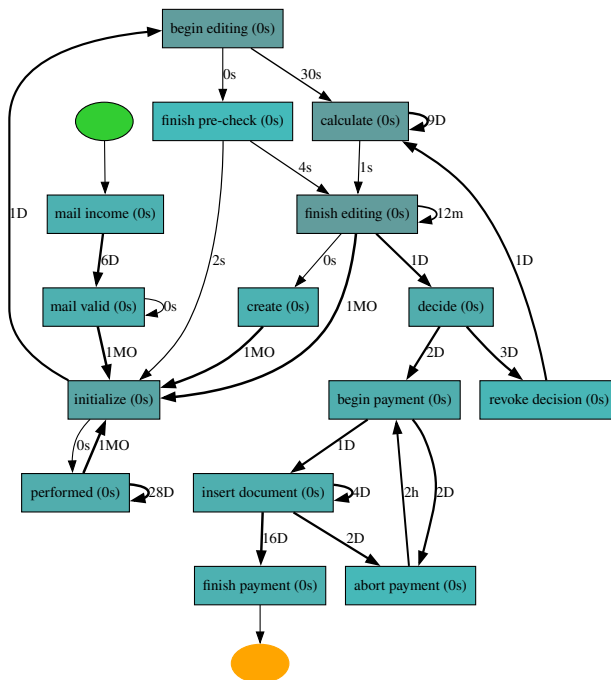


Figure 8.13: Heuristic Net - Top 10 variants - Dependency 0.80



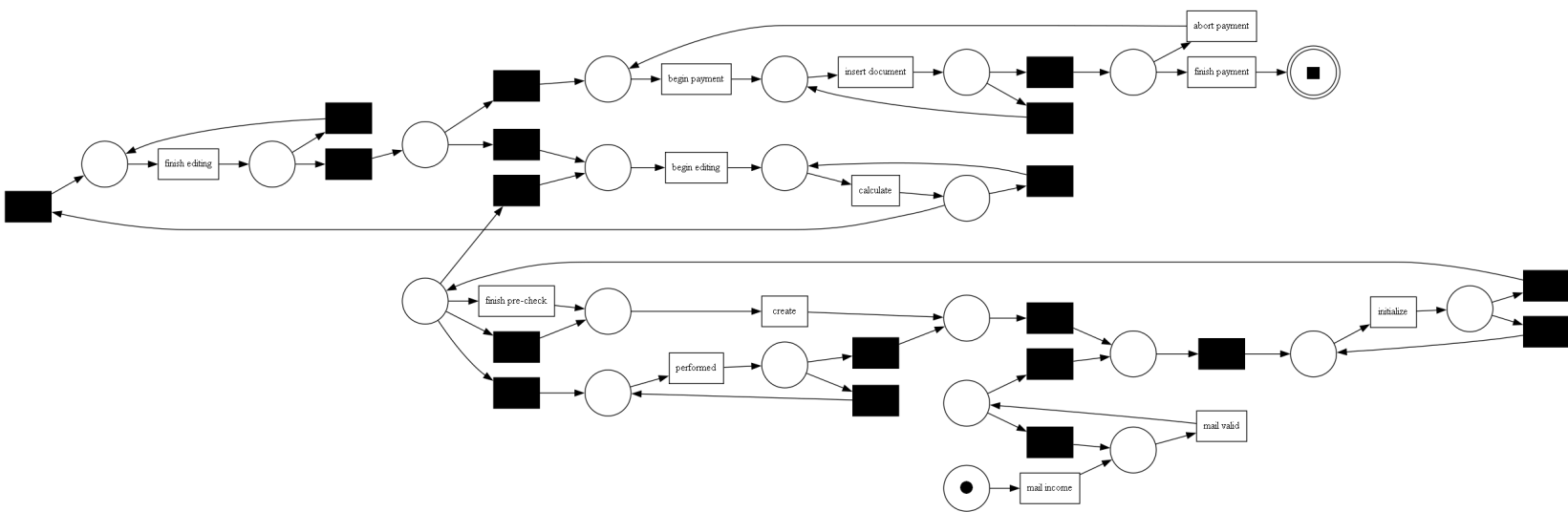


Figure 8.14: Inductive Miner - PetriNet - Reduced noise by 40%

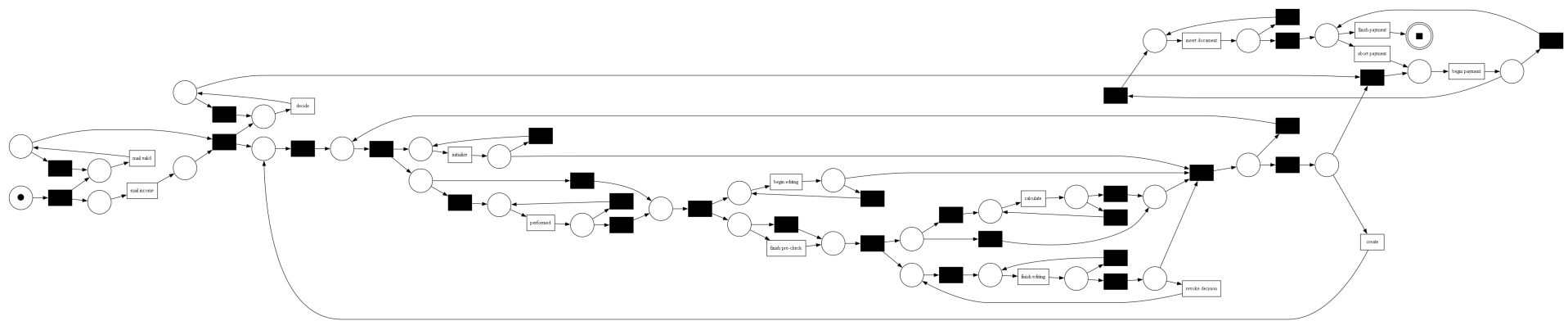


Figure 8.15: Inductive Miner - PetriNet - Top Ten Variants

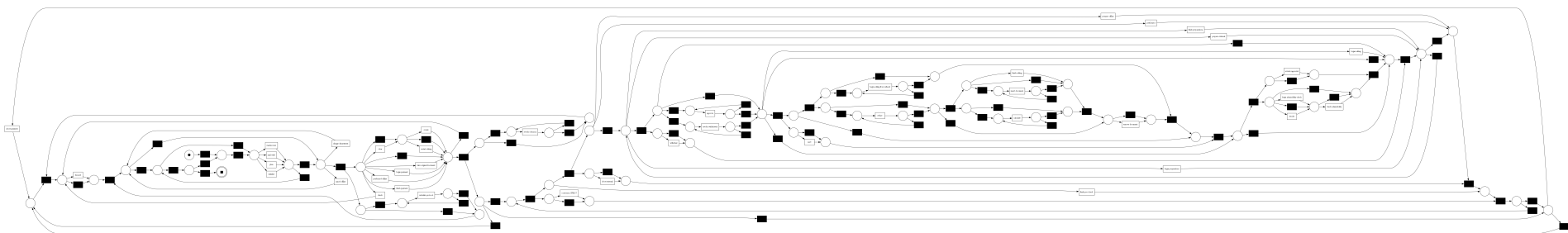


Figure 8.16: Inductive Miner - PetriNet - Total Event Log





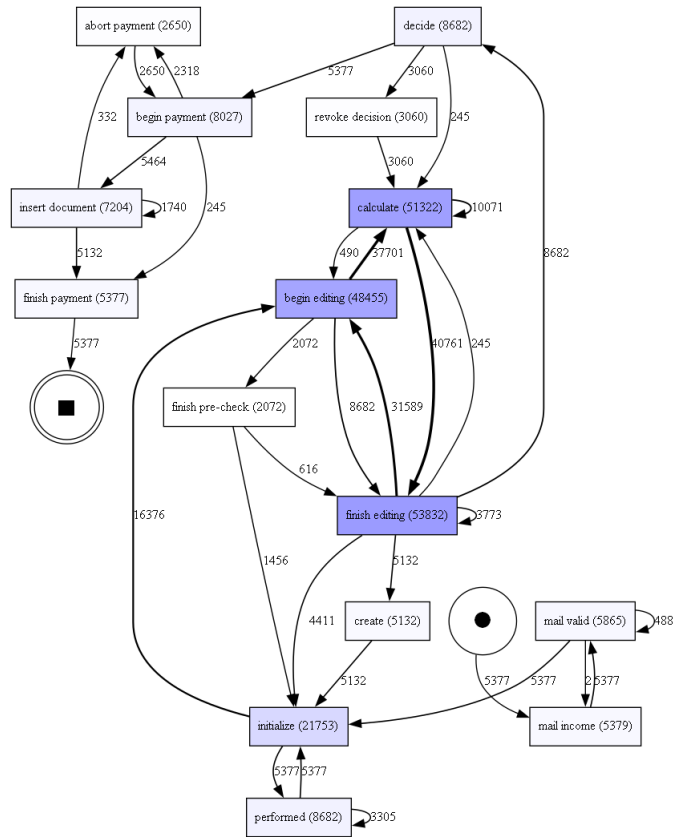


Figure 8.20: Directly Follows Graph - Top 10 Variants - Frequency Annotated

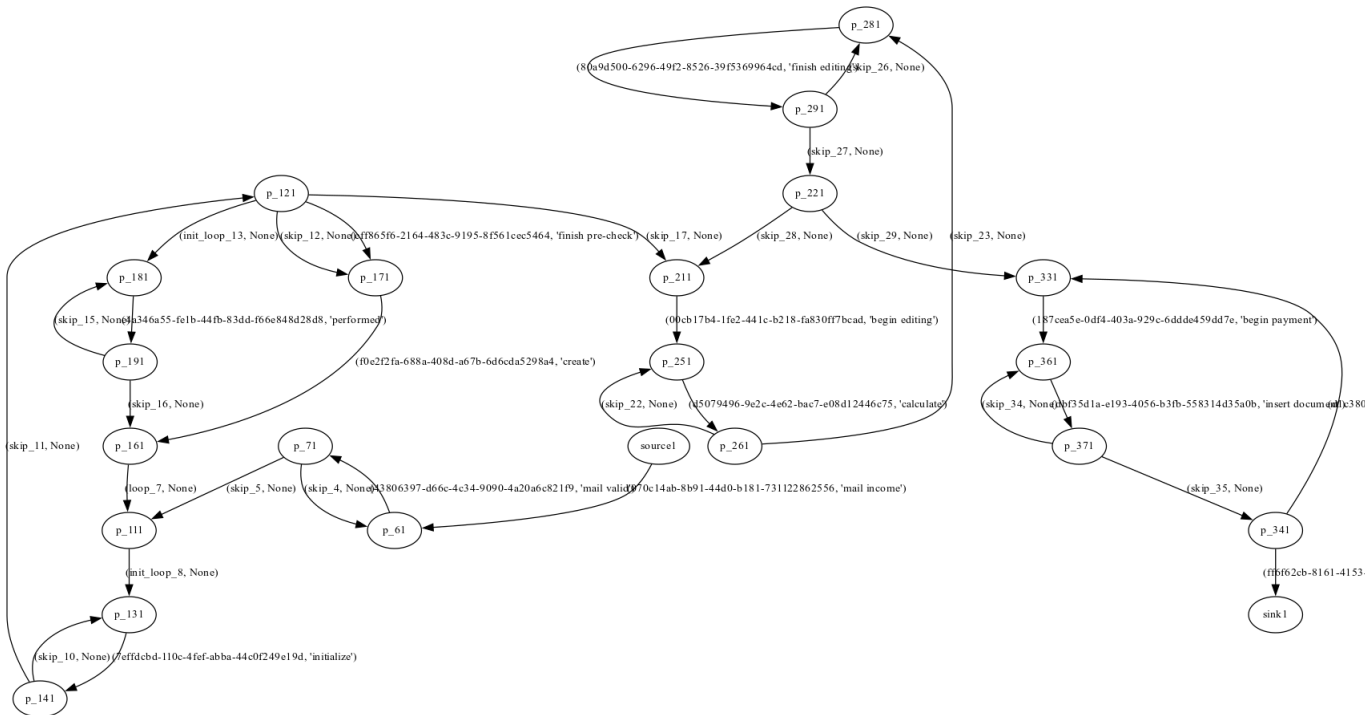


Figure 8.21: Reachability Graph - Inductive Miner

### 8.3 Έλεγχος Συμμόρφωσης

Ο έλεγχος συμμόρφωσης, όπως έχει αναφερθεί στην βιβλιογραφία, είναι μια μέθοδος σύγκρισης της εκτέλεσης μιας επιχειρηματικής διαδικασίας με το αντίστοιχο μοντέλο επιχειρηματικής διαδικασίας και κυρίως χρησιμοποιείται για την αξιολόγηση της πραγματικής συμπεριφοράς της διαδικασίας και τον εντοπισμό τυχόν διαφορών μεταξύ τους.

Αρχικά, γίνεται η εξέταση της σημειογραφίας PetriNet που προέκυψε από την εκτέλεση του αλγόριθμου "Heuristic Miner" [εικόνα 8.11]. Εφαρμόστηκαν δύο από τις τρεις τεχνικές ελέγχου συμμόρφωσης εξαιτίας του μοντέλου που παράγεται με την εφαρμογή του αλγόριθμου και τις όποιες ιδιαιτερότητες έχει.

Η αρχή γίνεται με την τεχνική του "Token Replay" κατά την οποία προέκυψαν τα παρακάτω αποτελέσματα:

- Από τα 5377 traces τα οποία αναπαράχθηκαν πάνω στο μοντέλο, στο τέλος της δεν ταίριαζε κανένα(0%).
- Ο μέσος όρος της ικανότητας επανάληψης ανά περίπτωση, δηλαδή το ποσοστό της περίπτωσης που γίνεται αποδεκτή από το μοντέλο διαδικασίας, είναι στο 0.92(92%).[8.6]
- Η συνολική καταλληλότητα(fitness) του αρχείου καταγραφής είναι επίσης στο 0.92(92%).
- Συνολικά αναπαρήχθηκαν 5377 περιπτώσεις με 26 καταστάσεις(place) και 15 μεταβάσεις(transitions). Δεν βρέθηκε καμία δραστηριότητα που να μην εκτελέστηκε.

Ο πίνακας 8.9 παρουσιάζει ενδεικτικά τις 50 πρώτες περιπτώσεις χρήσης που ελέγχθησαν με την τεχνική "Token Replay". Κάθε περίπτωση συνδέεται με το μοναδικό ID και τα αποτελέσματα δείχνουν αν ταίριαζε ή όχι, τα tokens που παρήχθησαν στο μοντέλο, αυτά που καταναλώθηκαν, τα missing και τα remaining tokens.

Table 8.6: Αποτελέσματα ελέγχου συμμόρφωσης του μοντέλου διαδικασιών[8.11] και του αρχείου καταγραφής γεγονότων με την τεχνική "Token Replay".

perc fit traces	average trace fitness	log fitness	percentage of fitting traces
0.0	0.9190647540870636	0.9205754490026536	0.0

Αντίθετα, όταν εφαρμόστηκε η τεχνική των ευθυγραμμίσεων (alignments) στο συγκεκριμένο μοντέλο η βιβλιοθήκη της pm4py έβγαλε Exception ότι δεν μπορεί να το διαχειριστεί λόγω του θορύβου(WARNING | The Petri net is not an easy sound net. Alignment cannot be applied. ).

Εφαρμόζοντας την τεχνική των footprints, όπως αυτή περιγράφεται στην θεωρία, γίνεται σύγκριση του footprint που υπολογίστηκε για το μοντέλο το οποίο εμφανίζεται στην εικόνα 8.22 με κάθε ένα από τα 10 διαφορετικά footprints που προκύπτουν από τα variants του τμήματος του event log που εξετάζουμε. Ο πίνακας 8.15 παρουσιάζει ενδεικτικά τις 50 πρώτες περιπτώσεις χρήσης της διαδικασίας και την ανάλυση τους. Σε γενικές γραμμές όσον αφορά το μοντέλο που ανακαλύφθηκε από τον Ευρετικό αλγόριθμο μέσα από τα αποτελέσματα του ελέγχου συμμόρφωσής βγαίνει το συμπέρασμα

ότι καμία απολύτως περίπτωση χρήσης δεν μπορεί να αναπαραχθεί πλήρως και αυτό οφείλεται στους περιορισμούς τους οποίους έχει η συγκεκριμένη τεχνική. Επίσης το γεγονός ότι έχει περιέχει σημεία τα οποία δεν είναι προσβάσιμα από την διαδικασία ή το γεγονός ότι περιέχει αδιέξοδα δεν μπορεί να γίνει η εφαρμογή της τεχνικής ελέγχου συμμόρφωσής που δίνει το καλύτερο αποτέλεσμα. Στην συνέχεια γίνεται εφαρμογή των συγκεκριμένων τεχνικών στο αντίστοιχο μοντέλο που προέκυψε από τον αλγόριθμο Inductive Miner.

Η αρχή γίνεται με την τεχνική του "Token Replay" κατά την οποία προέκυψαν τα παρακάτω αποτελέσματα,

- Από τα 5377 traces τα οποία αναπαράχθηκαν πάνω στο μοντέλο όλα αναπαράχθησαν με επιτυχία 1.0(100%).
- Ο μέσος όρος της ικανότητας επανάληψης ανά περίπτωση, δηλαδή το ποσοστό της περίπτωσης που γίνεται αποδεκτή από το μοντέλο διαδικασίας είναι στο 1.0(100%).[8.6]
- Η συνολική καταλληλότητα(fitness) του αρχείου καταγραφής είναι επίσης στο 1.0(100%).
- Συνολικά αναπαράχθηκαν 5377 περιπτώσεις με 35 καταστάσεις(place) να και 15 μεταβάσεις(transitions). Δεν βρέθηκε καμία δραστηριότητα που να μην εκτελέστηκε.

Ο πίνακας 8.9 παρουσιάζει ενδεικτικά τις 50 πρώτες περιπτώσεις χρήσης που ελέγχθησαν με την τεχνική "Token Replay". Κάθε περίπτωση συνδέεται με το μοναδικό ID και τα αποτελέσματα δείχνουν αν ταίριαζε ή όχι, τα tokens που παρήχθησαν στο μοντέλο, αυτά που καταναλώθηκαν, τα missing και τα remaining tokens. Βγαίνει το συμπέρασμα ότι το μοντέλο διαδικασιών που προέρχεται από τον Inductive Miner μπορεί να αναπαράγει πλήρως κάθε περίπτωση χρήσης που θα εφαρμοστεί σε αυτό. Ενδεικτικά βλέπουμε ότι για κάθε περίπτωση δεν έχει μείνει κανένα token από αυτά που έχουν παραχθεί. Αυτό συμβαίνει γιατί το Process Tree το οποίο είναι από κατασκευής deadlock-free και το μοντέλο που έχει παραχθεί είναι αυτό που παρουσιάζει τι πραγματικά συμβαίνει. Επίσης εφαρμόζοντας την τεχνική των alignments, επιβεβαιώνει ότι το μοντέλο το οποίο έχει προκύψει από τον αλγόριθμο και μπορεί να αναπαράγει πλήρως την διαδικασία. Τα αποτελέσματα των ευθυγραμμίσεων που προκύπτουν ανά διαδρομή παρουσιάζονται στην εικόνα[8.23]. Επίσης ο πίνακας 8.12 παρουσιάζει ενδεικτικά κάποια δείγματα αποτελεσμάτων του ελέγχου συμμόρφωσης ανά περίπτωση χρήσης και δείχνουν πως λειτουργεί η συγκεκριμένη τεχνική. Τελευταία τεχνική που χρησιμοποιείται για να πραγματοποιήσει έλεγχο συμμόρφωσης για το συγκεκριμένο μοντέλο είναι τα footprints. Η εικόνα 8.15 μας παρουσιάζει το footprint του μοντέλου διαδικασίας και ένα δείγμα από την σύγκριση που γίνεται, φαίνεται στον πίνακα 8.13. Επιβεβαιώνει δηλαδή το γεγονός ότι το μοντέλο διαδικασίας που έχει παραχθεί από τον Inductive Miner αντιπροσωπεύει την πραγματική διαδικασία.

Table 8.7: Αποτελέσματα ελέγχου συμμόρφωσης του μοντέλου διαδικασιών[8.15] και του αρχείου καταγραφής γεγονότων με την τεχνική "Token Replay".

perc fit traces	average trace fitness	log fitness	percentage of fitting traces
100.0	1.0	1.0	100.0



Table 8.8: Αποτελέσματα ελέγχου συμμόρφωσης του μοντέλου διαδικασιών[8.15] και του αρχείου καταγραφής γεγονότων με την τεχνική "Alignments".

<b>percFitTraces</b>	<b>averageFitness</b>	<b>log fitness</b>	<b>percentage of fitting traces</b>	<b>average trace fitness</b>
100.0	1.0	0.999876694406296	100.0	1.0

Table 8.9: Αποτελέσματα ελέγχου συμμόρφωσης (50 πρώτες περιπτώσεις) με την τεχνική αξιολόγησης token replay για το μοντέλο, εικόνα [8.11].

CaseID	Trace fit	is	Missing Tokens	Consumed Tokens	Remaining Tokens	Produced Tokens
d67b7d3aa510c7cd	False		6	92	8	94
065149cba1958a69	False		6	84	9	87
4676e26a64c8be2b	False		6	108	8	110
0c88e39e2a99d73e	False		6	108	8	110
2145469d8ced39f1	False		6	84	9	87
fadbe262528fb1ed	False		6	108	8	110
f90b9c51f743f687	False		7	85	11	89
8e2b67e54f182809	False		6	108	8	110
5e8f18aca6b44b37	False		6	92	8	94
c2516d9910a9f6f2	False		6	84	9	87
411f161ce480e774	False		6	108	8	110
fbf868b1e7698192	False		6	108	8	110
996cb528548cfc96	False		6	108	9	111
948fbd00ddb4719f	False		7	85	11	89
1c932dcc80c6c2bb	False		6	108	8	110
77eced495730c94c	False		6	81	9	84
0edb0cdbbbd94b55	False		7	93	11	97
3d9b10287b85830c	False		6	108	8	110
6d2a6a3294481944	False		6	81	9	84
44110e2f514e88cc	False		6	108	8	110
4342f6c989fe3e42	False		6	84	9	87
9d45dd35b16cc2e1	False		6	108	8	110
d7b45d16e7a6b1e1	False		6	84	9	87
4ff83fbe778f41bf	False		6	81	9	84
ddc17a805665b527	False		6	84	9	87
1afaaa0c20aee729	False		6	108	8	110
15aaced81f80b13d	False		6	92	8	94
69aa0a92070cf831	False		6	84	9	87
676eb052ead674e8	False		6	81	9	84
b5c35b73d3aa2bb0	False		6	119	8	121
74a8e60ac1fd2832	False		6	84	9	87
6c9425591896df59	False		6	81	9	84
b42ffcc9d10e812	False		7	93	11	97
05435ed96b738e20	False		6	81	9	84
4866f29be1f39624	False		7	85	11	89
8bb43904727f0cdb	False		6	84	9	87
ba014883f6eacd1e	False		6	119	8	121
c5c3f71222897f43	False		6	81	9	84
5380692ed64c229c	False		6	108	8	110
5a13746ab5e4c81e	False		6	108	8	110
7fa4657342a19dda	False		6	84	9	87
e55f7ead284dfbcf	False		6	108	8	110
c6bee0770ea5bdbf	False		6	108	8	110
69e4db24f9c0a162	False		6	108	8	110
3dd61a4848edd93e	False		6	81	9	84
066b7a1e5b2b5c87	False		7	93	11	97
a6e420f227d3532e	False		6	81	9	84
e10a2ba8f6cf5b2c	False		6	108	8	110
f615600f19b2c791	False		6	108	9	111
277d814bb5badb3b	False		6	81	9	84
706030bdd207ba0e	False		6	92	8	94
73b7197faadce113	False		6	119	8	121

Table 8.10: Αποτελέσματα ελέγχου συμμόρφωσης (50 πρώτες περιπτώσεις) με την τεχνική αξιολόγησης των footprints για το μοντέλο της εικόνα [8.11].

footprints ID	start activities	end activities	activities always happens	min length fit	is footprints fit
d67b7d3aa510c7cd	set()	set()	set()	True	False
065149cba1958a69	set()	set()	set()	True	False
4676e26a64c8be2b	set()	set()	set()	True	False
0c88e39e2a99d73e	set()	set()	set()	True	False
2145469d8ced39f1	set()	set()	set()	True	False
fadbe262528fb1ed	set()	set()	set()	True	False
f90b9c51f743f687	set()	set()	set()	True	False
8e2b67e54f182809	set()	set()	set()	True	False
5e8f18aca6b44b37	set()	set()	set()	True	False
c2516d9910a9f6f2	set()	set()	set()	True	False
411f161ce480e774	'mailvalid'	set()	set()	True	False
fbf868b1e7698192	set()	set()	set()	True	False
996cb528548cfc96	set()	set()	set()	True	False
948fbd00ddb4719f	set()	set()	set()	True	False
1c932dcc80c6c2bb	set()	set()	set()	True	False
77eced495730c94c	set()	set()	set()	True	False
0edb0cdbbbd94b55	set()	set()	set()	True	False
3d9b10287b85830c	set()	set()	set()	True	False
6d2a6a3294481944	set()	set()	set()	True	False
44110e2f514e88cc	set()	set()	set()	True	False
4342f6c989fe3e42	set()	set()	set()	True	False
9d45dd35b16cc2e1	set()	set()	set()	True	False
d7b45d16e7a6b1e1	set()	set()	set()	True	False
4ff83fbe778f41bf	'mailvalid'	set()	set()	True	False
ddc17a805665b527	set()	set()	set()	True	False
1afaaa0c20aee729	set()	set()	set()	True	False
15aaced81f80b13d	set()	set()	set()	True	False
69aa0a92070cf831	set()	set()	set()	True	False
676eb052ead674e8	set()	set()	set()	True	False
b5c35b73d3aa2bb0	set()	set()	set()	True	False
74a8e60ac1fd2832	set()	set()	set()	True	False
6c9425591896df59	set()	set()	set()	True	False
b42ffcc9d10e812	set()	set()	set()	True	False
05435ed96b738e20	'mailvalid'	set()	set()	True	False
4866f29be1f39624	set()	set()	set()	True	False
8bb43904727f0cdb	set()	set()	set()	True	False
ba014883f6eacd1e	set()	set()	set()	True	False
c5c3f71222897f43	set()	set()	set()	True	False
5380692ed64c229c	set()	set()	set()	True	False
5a13746ab5e4c81e	set()	set()	set()	True	False
7fa4657342a19dda	set()	set()	set()	True	False
e55f7ead284dfbcf	set()	set()	set()	True	False
c6bee0770ea5bdbf	set()	set()	set()	True	False
69e4db24f9c0a162	set()	set()	set()	True	False
3dd61a4848edd93e	set()	set()	set()	True	False
066b7a1e5b2b5c87	set()	set()	set()	True	False
a6e420f227d3532e	set()	set()	set()	True	False
e10a2ba8f6cf5b2c	'mailvalid'	set()	set()	True	False
277d814bb5badb3b	set()	set()	set()	True	False
706030bdd207ba0e	set()	set()	set()	True	False
73b7197faadce113	set()	set()	set()	True	False



Table 8.11: Αποτελέσματα ελέγχου συμμόρφωσης (50 πρώτες περιπτώσεις) με την τεχνική αξιολόγησης token replay για το μοντέλο, εικόνα [8.15]

CaseID	Trace fit	is	Missing Tokens	Consumed Tokens	Remaining Tokens	Produced Tokens
d67b7d3aa510c7cd	True		0	109	0	109
065149cba1958a69	True		0	116	0	116
4676e26a64c8be2b	True		0	121	0	121
0c88e39e2a99d73e	True		0	121	0	121
2145469d8ced39f1	True		0	116	0	116
fadbe262528fb1ed	True		0	121	0	121
f90b9c51f743f687	True		0	118	0	118
8e2b67e54f182809	True		0	121	0	121
5e8f18aca6b44b37	True		0	109	0	109
c2516d9910a9f6f2	True		0	116	0	116
411f161ce480e774	True		0	121	0	121
fbf868b1e7698192	True		0	121	0	121
996cb528548cfc96	True		0	131	0	131
948fbd00ddb4719f	True		0	118	0	118
1c932dcc80c6c2bb	True		0	121	0	121
77eced495730c94c	True		0	113	0	113
0edb0cdbbbd94b55	True		0	124	0	124
3d9b10287b85830c	True		0	121	0	121
6d2a6a3294481944	True		0	113	0	113
44110e2f514e88cc	True		0	121	0	121
4342f6c989fe3e42	True		0	116	0	116
9d45dd35b16cc2e1	True		0	121	0	121
d7b45d16e7a6b1e1	True		0	116	0	116
4ff83fbe778f41bf	True		0	113	0	113
ddc17a805665b527	True		0	116	0	116
1afaaa0c20aee729	True		0	121	0	121
15aaccd81f80b13d	True		0	109	0	109
69aa0a92070cf831	True		0	116	0	116
676eb052ead674e8	True		0	113	0	113
b5c35b73d3aa2bb0	True		0	130	0	130
74a8e60ac1fd2832	True		0	116	0	116
6c9425591896df59	True		0	113	0	113
b42ffcc9d10e812	True		0	124	0	124
05435ed96b738e20	True		0	113	0	113
4866f29be1f39624	True		0	118	0	118
8bb43904727f0cdb	True		0	116	0	116
ba014883f6eacd1e	True		0	130	0	130
c5c3f71222897f43	True		0	113	0	113
5380692ed64c229c	True		0	121	0	121
5a13746ab5e4c81e	True		0	131	0	131
7fa4657342a19dda	True		0	116	0	116
e55f7ead284dfbcf	True		0	131	0	131
c6bee0770ea5bdbf	True		0	121	0	121
69e4db24f9c0a162	True		0	121	0	121
3dd61a4848edd93e	True		0	113	0	113
066b7a1e5b2b5c87	True		0	124	0	124
a6e420f227d3532e	True		0	113	0	113
e10a2ba8f6cf5b2c	True		0	121	0	121
f615600f19b2c791	True		0	131	0	131
277d814bb5badb3b	True		0	113	0	113
706030bdd207ba0e	True		0	109	0	109

Year	Nigeria
2000	...
2001	...
2002	...
2003	...
2004	...
2005	...
2006	...
2007	...
2008	...
2009	...
2010	...
2011	...
2012	...
2013	...
2014	...
2015	...
2016	...
2017	...
2018	...
2019	...
2020	...
2021	...
2022	...
2023	...
2024	...
2025	...
2026	...
2027	...
2028	...
2029	...
2030	...
2031	...
2032	...
2033	...
2034	...
2035	...
2036	...
2037	...
2038	...
2039	...
2040	...
2041	...
2042	...
2043	...
2044	...
2045	...
2046	...
2047	...
2048	...
2049	...
2050	...

Figure 8.23: Alignment of the model [8.15]

Table 8.12: Αποτελέσματα ελέγχου συμμόρφωσης (50 πρώτες περιπτώσεις) με την τεχνική αξιολόγησης alignments για το μοντέλο, εικόνα [8.15]

Case ID	Cost	Queued States	Traversed Arcs	LP Solved	Fitness	BWC
d67b7d3aa510c7cd	57	1719	2155	52	1.0	510015
065149cba1958a69	63	1296	1563	47	1.0	490015
4676e26a64c8be2b	63	2168	2651	70	1.0	570015
0c88e39e2a99d73e	63	2168	2651	70	1.0	570015
2145469d8ced39f1	63	1296	1563	47	1.0	490015
fadbe262528fb1ed	63	2168	2651	70	1.0	570015
f90b9c51f743f687	64	1878	2434	64	1.0	500015
8e2b67e54f182809	63	2168	2651	70	1.0	570015
5e8f18aca6b44b37	57	1719	2155	52	1.0	510015
c2516d9910a9f6f2	63	1296	1563	47	1.0	490015
411f161ce480e774	63	2168	2651	70	1.0	570015
fbf868b1e7698192	63	2168	2651	70	1.0	570015
996cb528548cfc96	70	2470	3070	49	1.0	570015
948fd00ddb4719f	64	1878	2434	64	1.0	500015
1c932dcc80c6c2bb	63	2168	2651	70	1.0	570015
77eced495730c94c	62	1867	2426	64	1.0	470015
0edb0cdbbbd94b55	67	2296	3095	45	1.0	530015
3d9b10287b85830c	63	2168	2651	70	1.0	570015
6d2a6a3294481944	62	1867	2426	64	1.0	470015
44110e2f514e88cc	63	2168	2651	70	1.0	570015
4342f6c989fe3e42	63	1296	1563	47	1.0	490015
9d45dd35b16cc2e1	63	2168	2651	70	1.0	570015
d7b45d16e7a6b1e1	63	1296	1563	47	1.0	490015
4ff83fbe778f41bf	62	1867	2426	64	1.0	470015
ddc17a805665b527	63	1296	1563	47	1.0	490015
1afaaa0c20aee729	63	2168	2651	70	1.0	570015
15aaced81f80b13d	57	1719	2155	52	1.0	510015
69aa0a92070cf831	63	1296	1563	47	1.0	490015
676eb052ead674e8	62	1867	2426	64	1.0	470015
b5c35b73d3aa2bb0	67	1880	2220	52	1.0	620015
74a8e60ac1fd2832	63	1296	1563	47	1.0	490015
6c9425591896df59	62	1867	2426	64	1.0	470015
b42ffcc9d10e812	67	2296	3095	45	1.0	530015
05435ed96b738e20	62	1867	2426	64	1.0	470015
4866f29be1f39624	64	1878	2434	64	1.0	500015
8bb43904727f0cdb	63	1296	1563	47	1.0	490015
ba014883f6eacd1e	67	1880	2220	52	1.0	620015
c5c3f71222897f43	62	1867	2426	64	1.0	470015
5380692ed64c229c	63	2168	2651	70	1.0	570015
5a13746ab5e4c81e	70	1729	2058	43	1.0	570015
7fa4657342a19dda	63	1296	1563	47	1.0	490015
e55f7ead284dfbcf	70	1729	2058	43	1.0	570015
c6bee0770ea5bdbf	63	2168	2651	70	1.0	570015
69e4db24f9c0a162	63	2168	2651	70	1.0	570015
3dd61a4848edd93e	62	1867	2426	64	1.0	470015
066b7a1e5b2b5c87	67	2296	3095	45	1.0	530015
a6e420f227d3532e	62	1867	2426	64	1.0	470015
e10a2ba8f6cf5b2c	63	2168	2651	70	1.0	570015
f615600f19b2c791	70	2470	3070	49	1.0	570015
277d814bb5badb3b	62	1867	2426	64	1.0	470015
706030bdd207ba0e	57	1719	2155	52	1.0	510015
73b7197faadce113	67	1880	2220	52	1.0	620015

	abort payment	begin editing	begin payment	calculate	create	decide	finish editing	finish payment	finish pre-check	initialize	insert document	mail income	mail valid	performed	revoke decision
abort payment	#	#		#	#	#	#	#	#	#	<	#	#	#	#
begin editing	#	#	#		#	#		#	>	<	#	#	#	#	#
begin payment		#	#	#	#	<	#	>	#	#	>	#	#	#	#
calculate	#		#		#	<		#	#	#	#	#	#	#	<
create	#	#	#	#	#	#	<	#	#	>	#	#	#	#	#
decide	#	#	>	>	#	#	<	#	#	#	#	#	#	#	>
finish editing	#		#		>	>		#	<	>	#	#	#	#	#
finish payment	#	#	<	#	#	#	#	#	#	#	<	#	#	#	#
finish pre-check	#	<	#	#	#	#	>	#	#	>	#	#	#	#	#
initialize	#	>	#	#	<	#	<	#	<	#	#	#	<		#
insert document	>	#	<	#	#	#	#	>	#	#		#	#	#	#
mail income	#	#	#	#	#	#	#	#	#	#	#	#	>	#	#
mail valid	#	#	#	#	#	#	#	#	#	>	#	<		#	#
performed	#	#	#	#	#	#	#	#	#		#	#	#		#
revoke decision	#	#	#	>	#	<	#	#	#	#	#	#	#	#	#

Figure 8.24: Footprint of the model [8.15]



Table 8.13: Αποτελέσματα ελέγχου συμμόρφωσης (50 πρώτες περιπτώσεις) με την τεχνική αξιολόγησης footprints για το μοντέλο, εικόνα [8.15]

footprint ID	start activities	end activities	activities always happens	min length fit	is footprints fit
d67b7d3aa510c7cd	set()	set()	set()	True	True
065149cba1958a69	set()	set()	set()	True	True
4676e26a64c8be2b	set()	set()	set()	True	True
0c88e39e2a99d73e	set()	set()	set()	True	True
2145469d8ced39f1	set()	set()	set()	True	True
fadbe262528fb1ed	set()	set()	set()	True	True
f90b9c51f743f687	set()	set()	set()	True	True
8e2b67e54f182809	set()	set()	set()	True	True
5e8f18aca6b44b37	set()	set()	set()	True	True
c2516d9910a9f6f2	set()	set()	set()	True	True
411f161ce480e774	set()	set()	set()	True	True
fbf868b1e7698192	set()	set()	set()	True	True
996cb528548cfc96	set()	set()	set()	True	True
948fbd00ddb4719f	set()	set()	set()	True	True
1c932dcc80c6c2bb	set()	set()	set()	True	True
77eced495730c94c	set()	set()	set()	True	True
0edb0cdbbbd94b55	set()	set()	set()	True	True
3d9b10287b85830c	set()	set()	set()	True	True
6d2a6a3294481944	set()	set()	set()	True	True
44110e2f514e88cc	set()	set()	set()	True	True
4342f6c989fe3e42	set()	set()	set()	True	True
9d45dd35b16cc2e1	set()	set()	set()	True	True
d7b45d16e7a6b1e1	set()	set()	set()	True	True
4ff83fbe778f41bf	set()	set()	set()	True	True
ddc17a805665b527	set()	set()	set()	True	True
1afaaa0c20aee729	set()	set()	set()	True	True
15aaccd81f80b13d	set()	set()	set()	True	True
69aa0a92070cf831	set()	set()	set()	True	True
676eb052ead674e8	set()	set()	set()	True	True
b5c35b73d3aa2bb0	set()	set()	set()	True	True
74a8e60ac1fd2832	set()	set()	set()	True	True
6c9425591896df59	set()	set()	set()	True	True
b42ffcc9d10e812	set()	set()	set()	True	True
05435ed96b738e20	set()	set()	set()	True	True
4866f29be1f39624	set()	set()	set()	True	True
8bb43904727f0cdb	set()	set()	set()	True	True
ba014883f6eacd1e	set()	set()	set()	True	True
c5c3f71222897f43	set()	set()	set()	True	True
5380692ed64c229c	set()	set()	set()	True	True
5a13746ab5e4c81e	set()	set()	set()	True	True
7fa4657342a19dda	set()	set()	set()	True	True
e55f7ead284dfbcf	set()	set()	set()	True	True
69e4db24f9c0a162	set()	set()	set()	True	True
3dd61a4848edd93e	set()	set()	set()	True	True
066b7a1e5b2b5c87	set()	set()	set()	True	True
a6e420f227d3532e	set()	set()	set()	True	True
e10a2ba8f6cf5b2c	set()	set()	set()	True	True
f615600f19b2c791	set()	set()	set()	True	True
277d814bb5badb3b	set()	set()	set()	True	True
706030bdd207ba0e	set()	set()	set()	True	True
73b7197faadce113	set()	set()	set()	True	True

## 8.4 Αξιολόγηση Μοντέλου

Η αξιολόγηση του μοντέλου, όπως αναφέρθηκε και στο βιβλιογραφικό μέρος, θέλει να απαντήσει στο ερώτημα αν το μοντέλο αντανακλά την πραγματικότητα. Για να φτάσουμε σε αυτή την παραδοχή πρέπει ένα μοντέλο να έχει την ικανότητα να αναπαριστά την παρατηρούμενη συμπεριφορά (fitness), να είναι κατανοητό (simplicity), να είναι ακριβές (precision) ώστε να αποφεύγονται συνθήκες στις οποίες το μοντέλο μπορεί να περιλαμβάνει πολλά ψευδή θετικά αποτελέσματα και να μην αναπαράγει ακριβώς την παρατηρούμενη συμπεριφορά (underfitting) και την ίδια στιγμή να μπορεί να είναι ευέλικτο σε νέες περιπτώσεις χρήσης (overfitting). Τελευταίο κριτήριο είναι η γενίκευση (generalization) που δείχνει το κατά πόσο το μοντέλο διαδικασίας μπορεί να γενικεύει πέρα από τα παρατηρούμενα δεδομένα γεγονότων. Στόχος είναι να βρεθεί ισορροπία μεταξύ αυτών των κριτηρίων.

Αρχικά για τον Ευρετικό αλγόριθμο με βάση τις τιμές που υπολογίστηκαν, το μοντέλο διαδικασίας αξιολογείται ως εξής. Η τιμή καταλληλότητας  $0.920575449002653$  υποδηλώνει τον βαθμό στον οποίο το μοντέλο ταιριάζει με τα παρατηρούμενα δεδομένα γεγονότων. Μια τιμή καταλληλότητας που είναι πιο κοντά στο 1,0 υποδεικνύει μια καλή αντιστοίχιση μεταξύ του μοντέλου και της πραγματικής εκτέλεσης της διαδικασίας. Η τιμή ακρίβειας  $1,0$  υποδηλώνει ότι το μοντέλο διαδικασίας αναπαράγει την παρατηρούμενη συμπεριφορά χωρίς κανένα λανθασμένο θετικό αποτέλεσμα. Αυτό υποδηλώνει ότι όλες οι δραστηριότητες και οι ακολουθίες που αντλούνται από το μοντέλο είναι σωστές με βάση τα δεδομένα γεγονότων. Η τιμή γενίκευσης  $0.9536324529322736$  αντανακλά την ικανότητα του μοντέλου διαδικασίας να γενικεύει πέρα από τα παρατηρούμενα δεδομένα γεγονότων. Μια υψηλότερη τιμή γενίκευσης υποδεικνύει ότι το μοντέλο αναπαριστά τα υποκείμενα μοτίβα με ακρίβεια και μπορεί να αναπαράγει τη συμπεριφορά της διαδικασίας ακόμη και για ανεπιτήρητες περιπτώσεις. Η τιμή απλότητας στο  $0.6470588235294118$  υποδηλώνει το επίπεδο πολυπλοκότητας, η χαμηλή τιμή απλότητας υποδεικνύει ένα πολύπλοκο μοντέλο με μεγαλύτερο αριθμό δραστηριοτήτων, αποφάσεων ή εξαρτήσεων, ενώ μια υψηλή τιμή υποδεικνύει ένα απλούστερο μοντέλο με λιγότερα στοιχεία. Για τις τιμές που ανακαλύφθηκαν στην αξιολόγηση που έγινε με τα footprints, ο βαθμός καταλληλότητάς δεν αντικατοπτρίζει την πραγματικότητα λόγω της παθογένειας των footprints να την υπολογίσουν με σωστό τρόπο και οι υπόλοιπες τιμές είναι σχεδόν ίδιες με μικρές αποκλίσεις.

Για τη αξιολόγηση του "Inductive Miner", η τιμή καταλληλότητας  $1.0$  υποδεικνύει ότι το μοντέλο ταιριάζει απόλυτα με την παρατηρούμενη συμπεριφορά. Αυτό υποδεικνύει ότι το μοντέλο αναπαράγει με ακρίβεια τη συμπεριφορά της διαδικασίας. Η τιμή ακρίβειας  $0.29503473821997517$  υποδεικνύει ότι το μοντέλο διαδικασίας έχει χαμηλή ακρίβεια και το μοντέλο μπορεί να περιλαμβάνει πολλά ψευδή θετικά αποτελέσματα και να μην αναπαράγει ακριβώς την παρατηρούμενη συμπεριφορά (underfitting). Η τιμή γενίκευσης  $0.9836324529322736$  υποδεικνύει ότι το μοντέλο μπορεί να αναπαράγει με ακρίβεια τη συμπεριφορά της διαδικασίας ακόμη και για μη παρατηρούμενες περιπτώσεις. Τέλος, Η τιμή απλότητας στο  $0.6470588235294118$  υποδηλώνει το επίπεδο πολυπλοκότητας και η χαμηλή της υποδεικνύει ένα πολύπλοκο μοντέλο με μεγαλύτερο αριθμό δραστηριοτήτων, αποφάσεων ή εξαρτήσεων.

Συνολικά, βάσει των παρεχόμενων τιμών, τα μοντέλα διαδικασίας φαίνεται να έχουν υψηλό βαθμό καταλληλότητας, η ακρίβεια ανάλογα με την χρήση του αλγόριθμου διαφέρει, η τιμή γενίκευσης υποδεικνύει ότι το μοντέλο μπορεί να γενικεύσει πέρα από τα παρατηρούμενα δεδομένα. Ωστόσο, η τιμή απλότητας υποδεικνύει ότι το μοντέλο

μπορεί να είναι σχετικά πολύπλοκο πράγμα που οφείλεται στην ιδιαιτερότητα της διαδικασίας και την πολυπλοκότητα του αρχείου καταγραφής γεγονότων.

Table 8.14: Αποτελέσματα ελέγχου συμμόρφωσης με την τεχνική "Token Replay" και αξιολόγησης του μοντέλου διαδικασιών [8.11] που παράγεται από τον αλγόριθμο "Heuristic Miner".

<b>Fitness</b>	<b>Precision</b>	<b>Generalization</b>	<b>Simplicity</b>
0.920575449002653	1.0	0.9536324529322736	0.6470588235294118

Table 8.15: Αποτελέσματα ελέγχου συμμόρφωσης με την τεχνική "FootPrints" και αξιολόγησης του μοντέλου διαδικασιών [8.11] που παράγεται από τον αλγόριθμο "Heuristic Miner".

<b>Fitness</b>	<b>Precision</b>	<b>Generalization</b>	<b>Simplicity</b>
0.035541195476575124	1.0	0.9622259120123994	0.6470588235294118

Table 8.16: Αποτελέσματα ελέγχου συμμόρφωσης με την τεχνική "Token Replay" και αξιολόγησης του μοντέλου διαδικασιών [8.15] που παράγεται με τον αλγόριθμο inductive miner.

<b>Fitness</b>	<b>Precision</b>	<b>Generalization</b>	<b>Simplicity</b>
1.0	0.29503473821997517	0.987272790925424	0.6666666666666666

Table 8.17: Αποτελέσματα ελέγχου συμμόρφωσης με την τεχνική "Alignment" και αξιολόγησης του μοντέλου διαδικασιών [8.15] που παράγεται με τον αλγόριθμο inductive miner.

<b>Fitness</b>	<b>Precision</b>	<b>Generalization</b>	<b>Simplicity</b>
0.999876694406296	0.29269750057975463	0.987272790925424	0.6666666666666666

Table 8.18: Αποτελέσματα ελέγχου συμμόρφωσης με την τεχνική "Footprints" και αξιολόγησης του μοντέλου διαδικασιών [8.15] που παράγεται με τον αλγόριθμο inductive miner.

<b>Fitness</b>	<b>Precision</b>	<b>Generalization</b>	<b>Simplicity</b>
1.0	1.0	0.987272790925424	0.6666666666666666

## 8.5 Εξόρυξη Κοινωνικών Δικτύων

Το αρχείο καταγραφής γεγονότων μπορεί και αποθηκεύει συγκεκριμένες πληροφορίες που αφορούν την επίδραση των resources στην διαδικασία. Βάσει αυτής της ιδιότητας, γίνεται υπολογισμός και εξόρυξη διαφόρων στατιστικών, πινάκων και δικτύων τα οποία μας βοηθούν στην κατανόηση της διαδικασίας. Αρχικά, γίνεται ο υπολογισμός του πίνακα αντιστοίχισης πόρων-δραστηριοτήτων (Resource-Activity Matrix) [8.19]

ο οποίος μας δείχνει πόσες φορές εκτελέστηκε μια δραστηριότητα και ποιο από τα resources που συμμετέχουν στην διαδικασία την εκτέλεσε. Παρέχει βασικές πληροφορίες και απαντάει στο ερώτημα τι κάνει ποιος κατά την διάρκειά εκτέλεσης της διαδικασίας. Ο πίνακας παρουσιάζει τα resources τα οποία αποτελούνται από τους εργαζομένους, οι οποίοι για λόγους GDPR παρουσιάζονται με κρυπτογραφημένο ID, και ένα σύνολο πληροφοριακών συστημάτων (Document processing automaton - DPA, Inspection service - IS, Notification automaton - NA, Parcel automaton - PA, Processing automaton - PRA, Reference alignment processor - RAP) που συμμετέχουν στην διαδικασία.

Στο resource-activity matrix πίνακα παρουσιάζεται η σχέση μεταξύ των δραστηριοτήτων και των resources, ωστόσο δεν παρουσιάζονται οι σχέσεις των μεταξύ των resources έτσι ώστε να αποτυπωθεί η μεταξύ τους συνεργασία και να οπτικοποιηθεί η ροή της διαδικασίας. Για να δημιουργηθεί ένα δίκτυο το οποίο αποτυπώνει την συνεργασία μεταξύ των πόρων του οργανισμού πρέπει αρχικά να υπολογιστεί ένας πίνακας, ο "hand-over work matrix" [8.25]. Ο πίνακας αυτός παρουσιάζει πόσο συχνά παραδίδεται η δραστηριότητα από το ένα τμήμα/άτομο/σύστημά σε ένα άλλο. Στην συνέχεια, γίνεται η κατασκευή του αντίστοιχου κοινωνικού δικτύου όπου κάθε κόμβος αναπαριστά ένα ένα τμήμα/άτομο/σύστημά και οι ακμές είναι οι μεταξύ τους σχέσεις. Επιπλέον, κάθε ακμή έχει και το αντίστοιχο βάρος το οποίο υποδεικνύει πόσο σημαντική είναι η σχέση αυτή. Στον γράφο που προκύπτει εφαρμόζονται μετρήσεις που δείχνουν ποια resources συμμετέχουν περισσότερο στην εκτέλεση της διαδικασίας. Το "Handover of Work Social Network" [8.26] που ανακαλύφθηκε από την επεξεργασία του πίνακα μετρά πόσες φορές ένα άτομο/τμήμα/σύστημα ακολουθείται από μία άλλη αντίστοιχη οντότητα κατά την εκτέλεση μιας επιχειρηματικής διαδικασίας. Εκτελώντας πράξεις της θεωρίας γραφημάτων πάνω στον γράφο που ανακαλύφθηκε υπολογίζονται οι 10 πιο κεντρικοί κόμβοι [8.20] του δικτύου.

Στην συνέχεια, παρουσιάζεται το "Working Together" Social NetWork [8.28] το οποίο υπολογίζει πόσες φορές δύο οντότητες συνεργάζονται για την επίλυση μιας δραστηριότητας σε μία περίπτωση συνεργασίας. Παρατηρείται από τις μετρήσεις του γράφου ότι στο επίκεντρό της διαδικασίας βρίσκονται τα πληροφοριακά συστήματα που χρησιμοποιούν οι χρήστες αλλά και οι άνθρωποι που χειρίζονται την διαδικασία. Οι παραπάνω μετρήσεις φαίνονται από τους υπολογισμούς του πίνακα 8.21 και τα βάρη των ακμών του γράφου στην εικόνα 8.29. Τέλος, ακόμη μία δυνατότητά της εξόρυξης διαδικασιών σε ότι αφορά την παρουσίαση της οργανωτικής οπτικής της διαδικασίας είναι η ομαδοποίηση των resources σε ρόλους. Οι ρόλοι εξάγονται βάσει των δραστηριοτήτων που εκτελούνται από τα resources. Στον πίνακα 8.22 γίνεται η αντιστοίχιση των δραστηριοτήτων με τα resources και η ανακάλυψη των ρόλων και ποιά άτομα/τμήματα/συστήματα τους αποτελούν.

Table 8.19: Activity Resource Matix.

	abrt paym	beg ed	beg paym	calc	crt	dec	fin ed	fin paym	fin pre- ch	init	ins doc	mail inc	mail vaL	perf	rev dec
00037f	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	29.0
0087cf	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	41.0
019209	0.0	1.0	0.0	1.0	0.0	2.0	1.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	160.0
023bb9	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	91.0
03b214	0.0	0.0	0.0	7.0	0.0	12.0	7.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	44.0
08e484	0.0	729.0	0.0	729.0	0.0	0.0	729.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
0;n/a	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	5377.0	5867.0	0.0	0.0
155add	0.0	0.0	0.0	0.0	0.0	1217.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
1bd3b5	0.0	0.0	0.0	7.0	0.0	0.0	7.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
28b6bf	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	49.0
2ac4ac	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	18.0
2bf205	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	9.0
2c546f	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	45.0
2ca0ae	0.0	0.0	0.0	0.0	0.0	39.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	47.0
2dc625	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	7.0
36c75c	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	18.0
39be3e	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	37.0
4298e3	0.0	0.0	0.0	0.0	0.0	477.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	96.0
439089	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	38.0
465290	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	39.0
478c4f	0.0	0.0	0.0	2.0	0.0	0.0	2.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	133.0
483029	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	9.0
4af6fb	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	149.0
4b9a7f	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	18.0
4fac6d	0.0	245.0	0.0	245.0	0.0	0.0	245.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
51d239	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	1.0
520882	0.0	0.0	0.0	0.0	0.0	1890.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
556da9	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	40.0
5c1b3c	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	12.0
5d9e09	0.0	0.0	0.0	0.0	0.0	0.0	1.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
5dd4ec	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	39.0
5e018b	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	43.0
60b0b8	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	1.0
6d6ae5	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	3.0
7078c7	0.0	0.0	0.0	0.0	0.0	557.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
727350	0.0	34685.0	0.0	48268.0	0.0	0.0	37707.0	0.0	0.0	0.0	0.0	0.0	0.0	3305.0	0.0
75992a	0.0	1.0	0.0	1.0	0.0	24.0	1.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	34.0
76f30e	0.0	0.0	0.0	0.0	0.0	3.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	59.0
79367e	0.0	0.0	0.0	0.0	0.0	6.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	55.0
7b9b55	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	5.0
7d12ba	0.0	1.0	0.0	8.0	0.0	0.0	8.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	130.0
7d4a25	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	116.0
7fb8a5	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	40.0
822fb7	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	16.0
8a1fba	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	72.0
8beb64	0.0	0.0	0.0	0.0	0.0	22.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	46.0
97d224	0.0	1059.0	0.0	1059.0	0.0	0.0	1059.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	87.0
9e088c	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	11.0
9e337f	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	107.0
DP-Z	2650.0	0.0	8027.0	0.0	0.0	0.0	0.0	5377.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
DPA	0.0	7695.0	0.0	0.0	5132.0	0.0	10011.0	0.0	2072.0	13317.0	0.0	0.0	0.0	0.0	0.0
IS	0.0	1.0	0.0	0.0	0.0	0.0	1.0	0.0	0.0	1.0	0.0	0.0	0.0	0.0	0.0
NA	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	7204.0	0.0	0.0	0.0	0.0
PA	0.0	3058.0	0.0	0.0	0.0	0.0	3058.0	0.0	0.0	3058.0	0.0	0.0	0.0	0.0	0.0
PRA	0.0	980.0	0.0	980.0	0.0	0.0	980.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
RAP	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	3305.0	0.0	0.0	0.0	3305.0	0.0
a5ae3f	0.0	0.0	0.0	0.0	0.0	646.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
a5c3dd	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	49.0
ad25fc	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	103.0
aea7c8	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	33.0
af1e1e	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	172.0
b90293	0.0	0.0	0.0	1.0	0.0	2055.0	1.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	122.0
ba978b	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	24.0
bfa0ec	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	1.0
d85681	0.0	0.0	0.0	0.0	0.0	1680.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	15.0
d8639c	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	39.0
dde669	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	60.0
dee71b	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	31.0
dfd1d4	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	48.0
f6f7f1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	20.0
f9fe07	0.0	0.0	0.0	0.0	0.0	12.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
fa6437	0.0	0.0	0.0	12.0	0.0	7.0	12.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	57.0
fb5fa8	0.0	0.0	0.0	0.0	0.0	32.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	47.0
fb55b	0.0	0.0	0.0	2.0	0.0	1.0	2.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	232.0
ffb56b	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	13.0
scheduler	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	2072.0	0.0	0.0	0.0	2072.0	0.0

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	AA	AB	AC	AD	AE	AF	AG	AH	AI			
1		00037f	0087cf	19209	029bb9	03b214	08e484	0;n/a	155add	1bd9b5	28b6bf	2ac4ac	2bf205	2c546f	2ca0ae	2dc625	36c75c	39be3e	4.30E+06	439089	465290	478c4f	483029	4af6fb	4b9a7f	4fac6d	51d239	520882	556da9	5c1b3c	5d9e09	5dd4ec	5e018b	60b0b8	6d6ae5			
2	00037f	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0		
3	0087cf	0	0	0	0	0.000372	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0		
4	19209	0	0	0.000372	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0		
5	029bb9	0	0	0	0	0.000186	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0		
6	03b214	0	0	0	0	0.001302	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0		
7	08e484	0	0	0	0	0	0.271155	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0		
8	0;n/a	0	0	0	0	0	0	1.091129	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0		
9	155add	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0		
10	1bd9b5	0	0	0	0	0	0	0	0	0.001302	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
11	28b6bf	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0		
12	2ac4ac	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0		
13	2bf205	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0		
14	2c546f	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0		
15	2ca0ae	0	0.000186	0	0	0	0	0	0	0	0	0	0	0	0.000186	0	0	0	0	0.000186	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.000558	0		
16	2dc625	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0		
17	36c75c	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
18	39be3e	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
19	4.30E+06	0	0	0.000744	0	0	0	0	0	0	0	0	0	0.000372	0	0	0	0	0.000186	0	0.000186	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0		
20	439089	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0		
21	465290	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0		
22	478c4f	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.000372	0	0	0	0	0	0	0	0	0	0	0	0	0	
23	483029	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0		
24	4af6fb	0	0	0	0	0	0	0	0	0.001116	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
25	4b9a7f	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
26	4fac6d	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
27	51d239	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.091129	0	0	0	0	0	0	0	0	0	0	0	
28	520882	0	0.007067	0	0.016552	0.007997	0	0	0	0	0	0	0.001674	0	0.007997	0	0.006881	0	0.006695	0	0	0	0	0	0	0	0	0	0.007439	0	0.007253	0.007439	0	0	0	0		
29	556da9	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0		
30	5c1b3c	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
31	5d9e09	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
32	5dd4ec	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
33	5e018b	0	0	0	0	0.000186	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
34	60b0b8	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
35	6d6ae5	0	0	0	0	0	0	0	0	0.000186	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
36	7078c7	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
37	727350	0	0	0.000186	0	0	0.135577	0	0.180584	0	0	0	0	0	0.007253	0	0	0	0.088711	0	0	0	0	0	0	0	0	0.351497	0	0	0	0	0	0	0	0	0	
38	75992a	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
39	76f30e	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
40	79367e	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
41	7b9b55	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
42	7d12ba	0	0	0	0	0	0	0	0	0.000186	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
43	7d4a25	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	

Figure 8.25: Work Handover Matrix - Top 10 variants

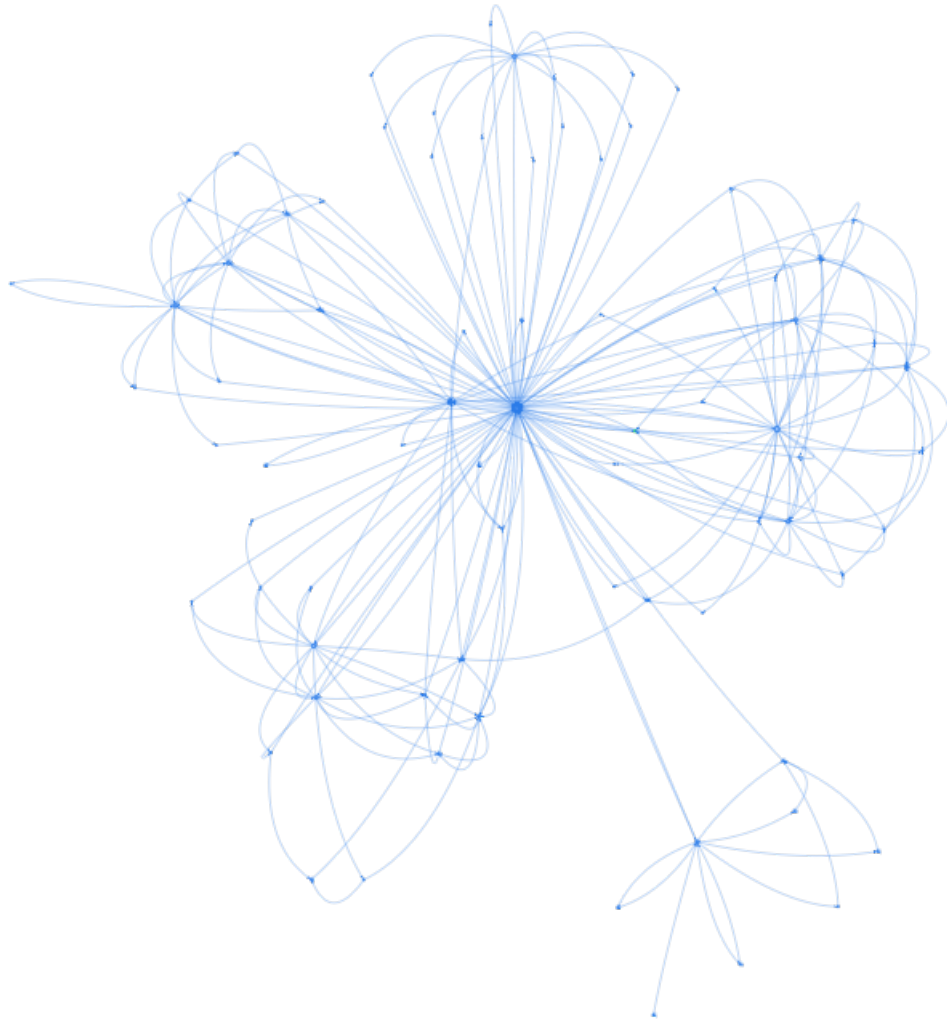
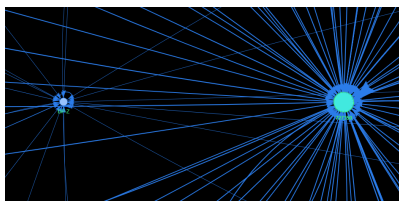
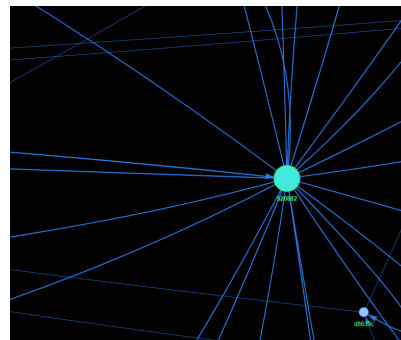


Figure 8.26: "Handover of Work" Social Network for Top 10 variants



(a) Resource:727350 and DP-Z



(b) Resource:520882

Figure 8.27: "Handover of Work" Social Network most Central Nodes

Table 8.20: Top ten nodes with the biggest Degree Centrality in Handover of Work Social Network.

Node	Node Degree	Degree Centrality
727350	66	0.2006079027355623
520882	23	0.06990881458966565
DP-Z	20	0.060790273556231005
a5ae3f	15	0.04559270516717325
d85681	15	0.04559270516717325
2ca0ae	15	0.04559270516717325
b90293	14	0.0425531914893617
fb5fa8	13	0.03951367781155015
4298e3	12	0.0364741641337386
8beb64	12	0.0364741641337386

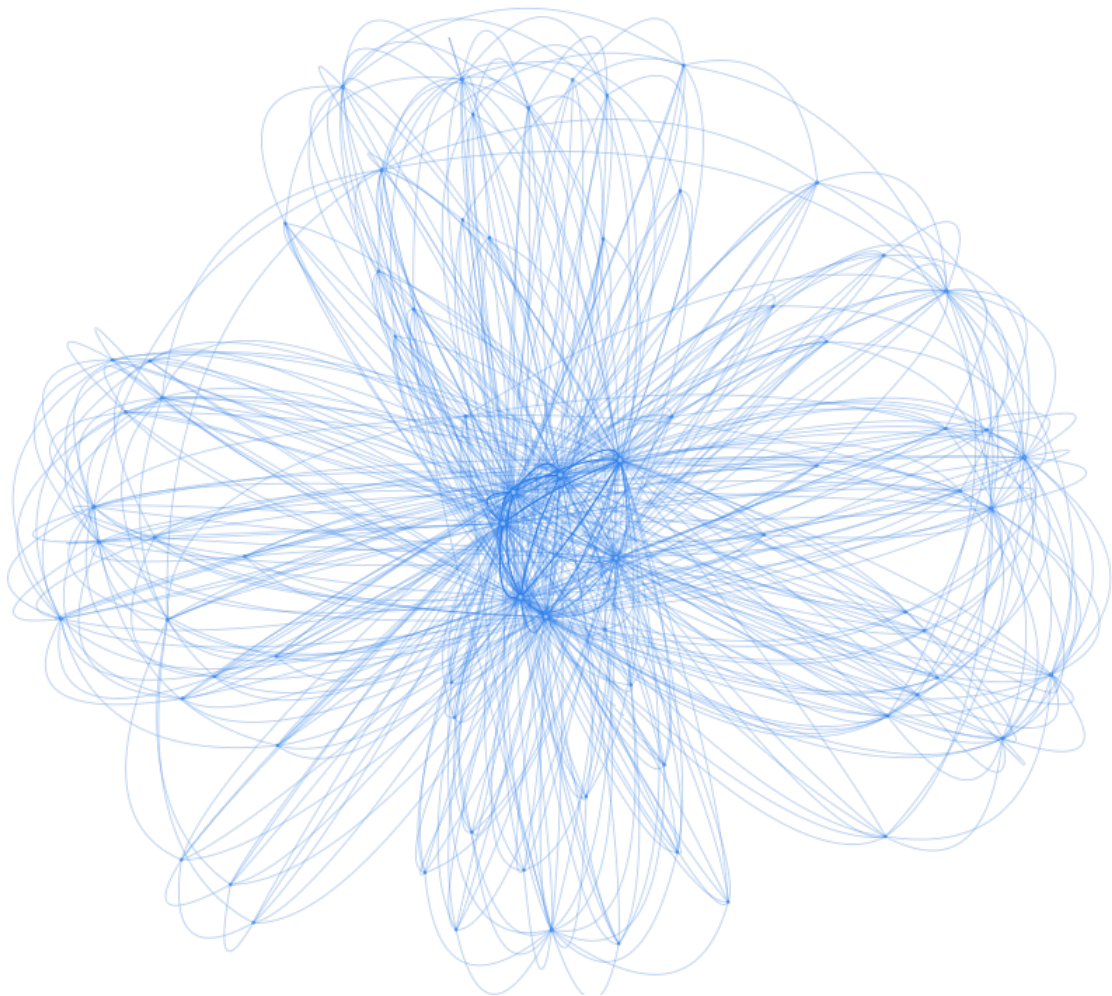


Figure 8.28: "Working Together" Social Network for Top 10 variants



Table 8.21: Top ten nodes with the biggest Degree Centrality in "Working Together" Social Network.

Node	Node Degree	Degree Centrality
0;n/a	75	0.05368647100930565
727350	75	0.05368647100930565
DP-Z	75	0.05368647100930565
Document processing automaton	75	0.05368647100930565
Notification automaton	72	0.051539012168933425
Reference alignment processor	71	0.050823192555476016
Parcel automaton	66	0.047244094488188976
520882	28	0.020042949176807445
2ca0ae	25	0.017895490336435216
fb5fa8	24	0.01717967072297781

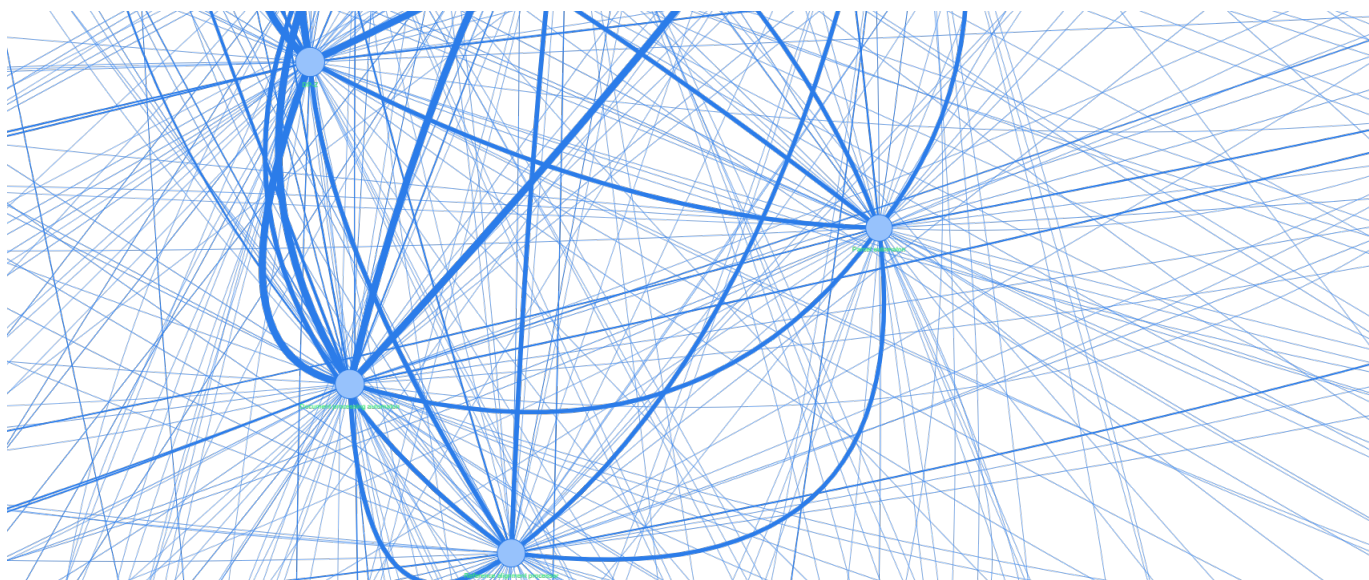


Figure 8.29: "Working Together" Nodes with high Centrality

Table 8.22: Roles.

Role	Activities	Number of Resources
0	['abort payment', 'begin payment', 'finish payment']	1
1	['begin editing', 'finish editing']	18
2	['calculate']	14
3	['create', 'finish pre-check']	1
4	['decide']	18
5	['initialize']	5
6	['insert document']	1
7	['mail income', 'mail valid']	1
8	['performed']	3
9	['revoke decision']	57

## 8.6 Ανάλυση Decision Point Mining

Τα μοντέλα διαδικασιών περιλαμβάνουν πολλά decision points, όπως για παράδειγμα, τα XOR-splits και τα XOR-joint. Με την χρήση των αρχείων καταγραφής και το μοντέλο διαδικασιών γίνεται προσπάθεια να συσχετιστεί η πραγματικότητα με το μοντέλο και σκοπός είναι κατανοηθεί γιατί προέκυψε μια συγκεκριμένη απόφαση και επιλέχτηκε ένα συγκεκριμένο μονοπάτι. Έτσι, εξαιτίας του όγκου των δεδομένων θα δούμε ένα μικρό σενάριο χρήσης αυτής της τεχνικής. Θα γίνει ταξινόμηση των περιπτώσεων χρήσης βάσει των μεταβλητών πρόβλεψης "young\_farmer" και "number\_parcel" σε σχέση με την εξαρτημένη μεταβλητή "penalty\_amount". Ουσιαστικά, προκύπτει το κατά πόσο επηρεάζει η ηλικία του αγρότη και ο αριθμός των αγροτεμαχίων που έχει το τελικό πόσο της αποζημίωσης. Το βασικό συμπέρασμα που προκύπτει βάσει της ανάλυσης αναδεικνύει ότι αγρότες κάτω από έναν αριθμό αγροτεμαχίων διέπονται σε διαφορετικούς κανόνες σχετικά με τις ποινές οι οποίοι δεν είναι τόσο αυστηροί και περνούν όλο το ποσό της αποζημίωσης, αντίθετα η ηλικία δεν παίζει τόσο ρόλο στο πώς υπολογίζονται οι ποινές. Επίσης, επιβεβαιώνονται αρχικοί ισχυρισμοί σχετικά με το αν εκτελείται σωστά η διαδικασία. Αιτούντες με μεγάλο αριθμό αγροτεμαχίων είναι πιο επιρρεπείς σε λάθη στην κατάθεση της αίτησης εξαιτίας και του διαφορετικού πλήθους εγγράφων που χρειάζονται.

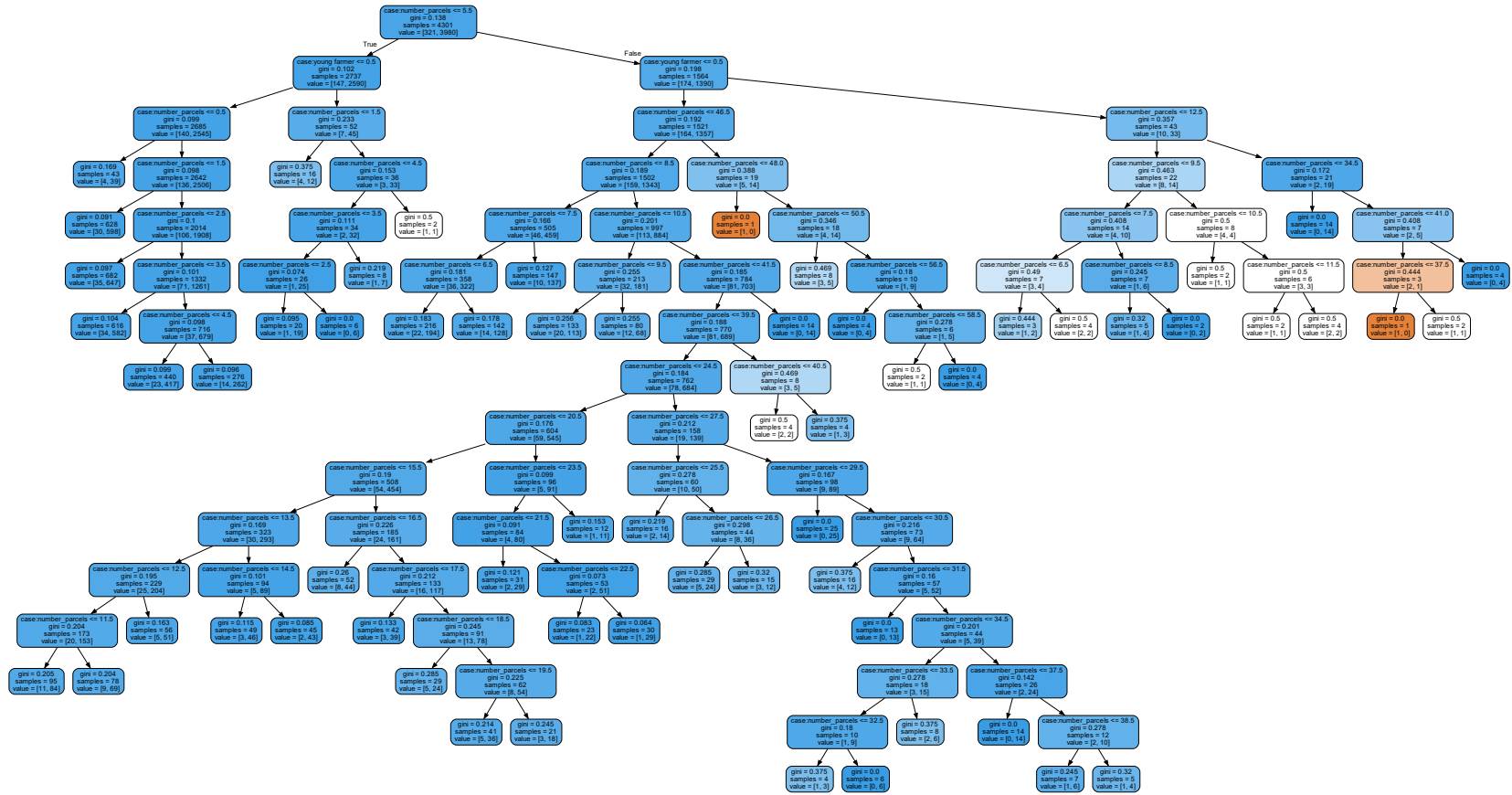


Figure 8.30: Decision Tree

# Συμπεράσματα και Μελλοντική Εργασία

## 9.1 Συμπεράσματα

Στην παρούσα διπλωματική εργασία πραγματοποιήθηκε ανάλυση και αξιολόγηση της διαδικασίας αίτησης άμεσων αποζημιώσεων από τον προϋπολογισμό του ταμείου Κοινής Αγροτικής Πολιτικής (CAP) της Ευρωπαϊκής Ένωσης. Ως κύριο σκοπό είχε να αναλύσει την διαδικασία, να δείξει πώς πραγματικά εκτελείται αντλώντας πληροφορίες μέσα από τα αρχεία καταγραφής γεγονότων και να διαπιστώσει πιθανά προβλήματά στον σχεδιασμό της και την εκτέλεση της από διαφορετικές οπτικές. Στο πλαίσιο της ανάλυσης και αξιολόγησης της συγκεκριμένης διαδικασίας, βγήκαν συμπεράσματα για σημεία που αφορούν την εξόρυξη διαδικασιών διευρύνοντας το αντικείμενο που εξετάζει η εργασία και αναλύονται σε βάθος επιπλέον ερωτήματα. Πρώτον, απαντάει στο ερώτημα γιατί σήμερα στην ανάλυση διαδικασιών, η οποία είναι μέρος ενός έργου διαχείρισης επιχειρηματικών διαδικασιών, είναι απαραίτητη η χρήση της τεχνικής του *Process Mining* και των εργαλείων που την συμπληρώνουν. Δεύτερον, παρουσιάζονται τα πλεονεκτήματα και τα μειονεκτήματα της χρήσης της Python σαν εργαλείο εξόρυξης διαδικασιών, τότε πρέπει να την χρησιμοποιούμε και τότε πρέπει να γίνει χρήση κάποιου άλλου εργαλείου.

Συνολικά, η διαδικασία αίτησης χρηματοδότησης είναι πολύπλοκη εκ φύσεως. Αυτό συμβαίνει εξαιτίας του πολύπλοκου εννοιολογικού πλαισίου εντός του οποίου λειτουργεί, δηλαδή ότι πρέπει ταυτόχρονα να συμμορφωθεί στους κανόνες που έχει θεσπίσει η Ευρωπαϊκή Ένωση και τους νόμους του κάθε κράτους ξεχωριστά. Επιπλέον, η διαδικασία είναι υβριδική, καθώς εμπλέκονται σε αυτή πληροφορικά συστήματα, οι χρήστες των συστημάτων, εξωτερικοί συνεργάτες που κάνουν τους έλεγχους και οι ίδιοι οι δικαιούχοι των αποζημιώσεων. Συνεπώς, ο ανθρώπινος παράγοντας ο οποίος συμμετέχει στην διαδικασία είναι απρόβλεπτος και έρχεται να επιβεβαιώσει την θεωρία της μοντελοποίησης μιας διαδικασίας, δηλαδή ότι η μη ικανότητά να καταγραφεί η ανθρώπινη συμπεριφορά επειδή επηρεάζεται από διάφορους παράγοντες, όπως οι ατομικές προτιμήσεις, οι γνωστικές παράμετροι, η κοινωνική δυναμική και το περιβαλλοντικό πλαίσιο. Αυτά τα στοιχεία μπορούν να εισάγουν ένα επίπεδο μη πρόβλεψης συμπεριφορών και πολυπλοκότητας που είναι δύσκολο να καταγραφεί μόνο μέσω μαθηματικών μοντέλων. Είναι σημαντικό να αναφερθεί ότι για τις 43809 περιπτώσεις η διαδικασία είχε 28457 διαφορετικές παραλλαγές από τις οποίες οι 26602 έχουν εκτελεστεί μοναδικά, γεγονός που κάνει δύσκολη την ανάλυση της διαδικασίας ως προς την ανίχνευση αποκλίσεων ή καθυστερήσεων κατά την εκτέλεσή της. Επίσης, αναλύοντας το σύνολο των δεδομένων και αρχικά λαμβάνοντας τους χρόνους έναρξης, τους

χρόνους ολοκλήρωσης αλλά και την περιγραφή της διαδικασίας, η εργασία καταλήγει στο συμπέρασμα ότι πρόκειται για μια διαδικασία η οποία άλλαζε με το πέρασμα των χρόνων. Η διαδικασία πραγματοποιείται και τις τρεις χρονιές για τις οποίες μας παρέχεται ιστορικό δεδομένων σχεδόν τα ίδια άτομα. Αυτό είχε αποτέλεσμα σε βάθος χρόνου οι αιτούντες και όσοι εμπλέκονται στην διαδικασία να προσαρμόζονται, να μαθαίνουν τον χειρισμό των συστημάτων και να κάνουν υποκειμενικές κρίσεις, πράγμα που προσθέτει εμπειρία και να εκτελείται η διαδικασία με τέτοιο τρόπο ώστε το χρηματικό όφελος να είναι μεγαλύτερο και να ολοκληρώνεται σε λιγότερο χρόνο και με λιγότερες δραστηριότητες. Επιπλέον, η βελτίωση της διαδικασίας οφείλεται στο γεγονός ότι ο ίδιος ο φορέας που καθορίζει την διαδικασία εντόπισε προβλήματα στην εκτέλεση της και προχώρησε σε συγκεκριμένες αλλαγές οι οποίες περιγράφονται στον πίνακα εγγράφων 5.1.

Από την ανάλυση που έγινε, εξήχθησαν επιπλέον συμπεράσματα σχετικά με την ποιότητα των δεδομένων που υπάρχουν στο αρχείο καταγραφής γεγονότων και από αυτά προέκυψε η ανακάλυψη της διαδικασίας. Αρχικά είχε κάποια μικρά λάθη που διορθώθηκαν κατά την διάρκεια της προ-επεξεργασίας, όπως για παράδειγμα λάθος ημερομηνίες σε επίπεδο έτους (άρχιζαν από το 2013 και 2014). Άλλα ένα σημαντικό θέμα, κατά την κρίση του γράφοντος, που επηρέασε και την ανακάλυψη των μοντέλων διαδικασιών είναι ότι οι δραστηριότητες με τις οποίες οι αλγόριθμοι ανακάλυψης διαδικασιών εξάγουν το αντίστοιχο μοντέλο πολλές φορές δεν ήταν διακριτές. Για παράδειγμα η δραστηριότητα 'save' είναι μέλος και του εγγράφου "Geo Parcel Document" αλλά και του "Reference alignment". Για αυτό το λόγο δημιουργήθηκε η ανάγκη να παρουσιαστεί η διαδικασία σε επίπεδο εγγράφων. Επίσης, υπήρχε η σκέψη να γίνει επιπλέον προ επεξεργασία δεδομένων ώστε οι δραστηριότητες να γίνουν διακριτές αλλά θα άλλαζε κατά πολύ το dataset και άρα, δεν θα προέκυπταν ασφαλή συμπεράσματα για την εργασία ανάλυσης. Πρόκειται για ένα dataset το οποίο παρουσιάζει θετικά χαρακτηριστικά ως προς την ποσότητά των δεδομένων και των ιδιοτήτων. Περιέχει δεδομένα τα οποία περιέχουν πληροφορίες έτσι ώστε να κατασκευαστεί το βέλτιστο μοντέλο διαδικασιών και να μην χρειάζονται επιπλέον στάδια ανάλυσης.

Αναφορικά με την χρήση της Python και της βιβλιοθήκης *pm4py* ως εργαλείο εξόρυξης διαδικασιών προκύπτουν επίσης ποιοτικά συμπεράσματα. Αρχικά, η χρήση της βιβλιοθήκης σε συνδυασμό με υπόλοιπες βιβλιοθήκες της γλώσσας προσφέρουν μεγάλη ευελιξία ως προς την εξαγωγή των αποτελεσμάτων που επιθυμεί ο κάθε χρήστης. Μπορεί, κάλλιστα, να χρησιμοποιηθεί πρόσθετα σε λογισμικό που βασίζεται στην Python και να εκτελέσει εργασίες εξόρυξης διαδικασιών. Τα αποτελέσματα που εξάγει είναι ορθά, τα μοντέλα διαδικασιών ευδιάκριτα και κατανοητά και στις τελευταίες εκδόσεις έχει αλλάξει την δομή των αρχείων καταγραφής δεδομένων από EventLog Class σε Dataframe με αποτέλεσμα να ενισχύει της ευελιξία της. Στα αρνητικά, είναι ότι με την διαδικασία η οποία αναλύθηκε, επιβεβαιώθηκε ότι, ειδικά σε προσωπικό υπολογιστή, έχει περιορισμένη απόδοση και η εκτέλεση των εργασιών είναι χρονοβόρα. Ο μέσος χρόνος επεξεργασίας των δεδομένων και η εκτέλεση του Python Script που κάνει την συγκεκριμένη ανάλυση είναι στα 70 λεπτά. Σε μερικές περιπτώσεις κατά την διάρκεια των δοκιμών και εργασιών προ επεξεργασίας του αρχείου καταγραφής γεγονότων ο χρόνος εκτέλεσης της ανάλυσης έφτασε τις 8 ώρες με χρήση προσωπικού υπολογιστή. Επιπλέον, πρέπει κάποιος να είναι προσεκτικός με τις αλλαγές των εκδόσεων της βιβλιοθήκης. Σίγουρα βελτιώνεται αλλά θέλει προσοχή γιατί συχνά οι αλλαγές είναι καθολικές ή καμιά φορά μπαίνουν λειτουργίες ώστε να χρησιμοποιηθούν από την κοινότητα και σε περίπτωση αρνητικής ανατροφοδότησης αφαιρούνται. Χαρακτηριστικό

παράδειγμα, το οποίο χρησιμοποιήθηκε και στην ανάπτυξη της παρούσας εργασίας, είναι η χρήση αποστολής μηνυμάτων προς την πλατφόρμα της OpenAi ChatGPT<sup>1</sup> η οποία προστέθηκε στην έκδοση 2.7.2, άλλαξε στην 2.7.4 και αφαιρέθηκε τελείως στην 2.7.5. Σε γενικές γραμμές, όμως, αποτελεί ένα ισχυρό εργαλείο για κάποιον που θέλει να πραγματοποιήσει αντίστοιχες εργασίες και είναι εξοικειωμένος με την Python.

Τέλος, τα οφέλη του Process Mining έχουν αναφερθεί κατά την διάρκεια του βιβλιογραφικού μέρους της εργασίας και αρκετά επιβεβαιώνονται στο ερευνητικό της μέρος. Σημαντικό είναι, ότι μπορούν να οπτικοποιήσουν την διαδικασία όσο πολύπλοκη και να είναι, ειδικά οι πιο προηγμένοι αλγόριθμοι ανακάλυψης διαδικασιών μπορούν να ανακαλύψουν μοντέλα που αναπαράγουν κάθε πιθανή παραλλαγή εκτέλεσης της ροής των δραστηριοτήτων. Και με τις επιπλέον παραλλαγές ανάλυσης όπως για παράδειγμα το decision point mining και την ανάλυση της κοινωνικής πτυχής της διαδικασίας ο ενδιαφερόμενος αποκτά απόλυτη γνώση σχετικά με το τί συμβαίνει, ποιος κάνει τί και γιατί επιλέγεται μια δραστηριότητα έναντι άλλης. Σε γενικές γραμμές, είναι μια μέθοδος που θα υιοθετείται από επιχειρήσεις και οργανισμούς όλο και περισσότερο εξαιτίας της ακρίβειας των αποτελεσμάτων της.

## 9.2 Μελλοντική Εργασία

Για την συγκεκριμένη διπλωματική εργασία δημιουργήθηκε πηγαίος κώδικας σε μορφή Script ο οποίος εκτελεί όλες τις εργασίες που την αποτελούν. Από την προ-επεξεργασία των δεδομένων μέχρι την ανακάλυψη των μοντέλων αξιολόγησης και κοινωνικών δικτύων σε μορφή γράφου. Η σκέψη για μελλοντική εργασία είναι αυτός ο κώδικας να αποτελέσει τη βάση για ανάπτυξη ενός εργαλείου το οποίο θα μπορεί να εκτυπώσει μια αναφορά η οποία θα δίνει λεπτομερή ανάλυση για την διαδικασία την οποία θα δέχεται σαν είσοδο με την μορφή πρότυπου IEEE XES ή σε μορφή .csv. Για να γίνει αυτό θα προστεθεί λειτουργικότητα γραφικής αλληλεπίδρασης χρήστη (GUI) και εφαρμογής με την οποία ο χρήστης θα μπορεί να επιλέγει το αρχείο καταγραφής γεγονότων που θέλει να επεξεργαστεί και να εφαρμόσει την κατάλληλη παραμετροποίηση. Στην παρούσα φάση τα αποτελέσματα εκτυπώνονται σε αρχείο .txt σε μορφή logs, τα plots, τα μοντέλα διαδικασίας και τα διαγράμματα διαδικασιών αποθηκεύονται με την μορφή .png στο τοπικό σύστημα φακέλων και οι πίνακες σε μορφή .csv. Η ιδέα είναι να εκτυπώνεται μία αναφορά σε μορφή .pdf με όλα τα αποτελέσματα που έχουν προκύψει. Όσον αφορά την ίδια τη διαδικασία, δεν χρίζει περαιτέρω επεξεργασίας καθώς με την πάροδο των χρόνων η Ευρωπαϊκή Ένωση, και συγκεκριμένα τα έτη 2020 και 2023, την έχει αλλάξει καθολικά επιβάλλοντας νέους κανόνες και θεσπίζοντας νέους νόμους οι οποίοι αφορούν την Κοινή Αγροτική Πολιτική. Μπορούμε να δούμε την εξέλιξη της διαδικασίας μέσα από τον ιστότοπο <sup>2</sup> που αναφέρει και παρουσιάζει την συγκεκριμένη διαδικασία και τις αλλαγές που έχουν προκύψει.

<sup>1</sup><https://chat.openai.com/>

<sup>2</sup>[https://agriculture.ec.europa.eu/common-agricultural-policy/cap-overview/cap-glance\\_en#timeline](https://agriculture.ec.europa.eu/common-agricultural-policy/cap-overview/cap-glance_en#timeline)







# Βιβλιογραφία

- [1] Omar AlShathry. Process mining as a business process discovery technique. *International Journal of Advanced Computer Science and Applications*, 10(3):149–157, 2019.
- [2] Joos C.A.M. Buijs and Wil M.P. Van der Aalst. Turning big data into value: A literature review on business value realization from process mining. *Information Systems*, 59:143–156, 2016.
- [3] Gustavo Callou, Paulo Maciel, Carlos Araújo, João Ferreira, and Rafael Souza. *A Petri Net-Based Approach to the Quantification of Data Center Dependability*, pages 313–336. CHAPTER METRICS OVERVIEW, 08 2012.
- [4] David Chapela-Campa, Marlon Dumas, Manuel Mucientes, and Manuel Lama. Efficient edge filtering of directly-follows graphs for process mining. *Information Sciences*, 610, 08 2022.
- [5] Data Experts. Data experts. <https://www.data-experts.de/>. Accessed on May 16, 2023.
- [6] Marlon Dumas, Marcello La Rosa, Jan Mendling, and Hajo A. Reijers. *Business Process Management: A Comprehensive Survey*. Springer Science & Business Media, New York, 1st edition, 2013.
- [7] European Commission - Agriculture and Rural Development. Basic payment. [https://agriculture.ec.europa.eu/common-agricultural-policy/income-support/basic-payment\\_en](https://agriculture.ec.europa.eu/common-agricultural-policy/income-support/basic-payment_en).
- [8] European Commission - Agriculture and Rural Development. Common agricultural policy. [https://agriculture.ec.europa.eu/common-agricultural-policy\\_en](https://agriculture.ec.europa.eu/common-agricultural-policy_en).
- [9] European Commission - Agriculture and Rural Development. European commission direct payments for farmers 2015-2020. [https://ec.europa.eu/agri-culture/sites/agriculture/files/direct-support/direct-payments/docs/\\_redirect.htm?lang=en](https://ec.europa.eu/agri-culture/sites/agriculture/files/direct-support/direct-payments/docs/_redirect.htm?lang=en).
- [10] Horses for Sources. Digital workforce to shine. [https://www.horsesforsources.com/digital-workforce-to-shine\\_041820](https://www.horsesforsources.com/digital-workforce-to-shine_041820), 2020. Accessed: 2022-03-25.
- [11] Grand View Research. Process mining software market report. <https://www.grandviewresearch.com/industry-analysis/process-mining-software-market-report>, 2023.

- [12] John C Henderson and N Venkatraman. Strategic alignment: Leveraging information technology for transforming organizations. *IBM systems journal*, 38(2.3):472–484, 1999.
- [13] Sarah Leemans and Dirk Fahland. Process mining: A 360 degree overview. In *Process Mining*, pages 1–33. Springer, 2019.
- [14] Jerry Luftman. Assessing business-it alignment maturity. *Communications of the Association for Information Systems*, 4, 2000.
- [15] Badakhshan Peyman, Wurm Bastian, Grisold Thomas, Geyer-Klingenberg Jerome, Mendling Jan, and vom Brocke Jan. Creating business value with process mining. *Journal of Strategic Information Systems*, 31(2):101745, 2022.
- [16] PM4Py Contributors. PM4Py: Process Mining for Python. <https://pm4py.fit.fraunhofer.de/>, 2023.
- [17] PR Newswire. Business process management (bpm) industry outlook to 2025 - growing need to optimize business processes for meeting the dynamic requirements of customers. <https://www.prnewswire.com.html>, 2020. Accessed: 2022-03-25.
- [18] John Smith. *Business Manager Fundamentals*. Acme Publishing, New York, 1st edition, 2020.
- [19] W. M. P. van der Aalst. *Process mining: discovery, conformance and enhancement of business processes*. Springer Science & Business Media, 2011.
- [20] Wil van der Aalst. *Foundations of Process Event Data*. Springer, Berlin, Heidelberg, 2016.
- [21] Wil van der Aalst. *Process Mining: Data Science in Action*. Springer, Cham, 2016.
- [22] Wil van der Aalst. *Creating Business Value with Process Mining*. Springer, Cham, 2018.
- [23] Wil M. P. van der Aalst, Ton Weijters, and Laura Maruster. Workflow mining: Discovering process models from event logs. *IEEE Transactions on Knowledge and Data Engineering*, 16(9):1128–1142, 2004.
- [24] W.M.P. van der Aalst and J. Carmona. *Process Mining Handbook*. Lecture Notes in Business Information Processing. Springer International Publishing, 2022.
- [25] Boudewijn van Dongen and F. (Florian) Borchert. Bpi challenge 2018, 2018.
- [26] Mathias Weske. *Business Process Management: Concepts, Languages, Architectures*. Springer, Berlin, Heidelberg, 2012.