



ΠΑΝΕΠΙΣΤΗΜΙΟ ΔΥΤΙΚΗΣ ΑΤΤΙΚΗΣ
ΣΧΟΛΗ ΜΗΧΑΝΙΚΩΝ
ΤΜΗΜΑ ΜΗΧΑΝΙΚΩΝ ΠΛΗΡΟΦΟΡΙΚΗΣ ΚΑΙ
ΥΠΟΛΟΓΙΣΤΩΝ

ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

Πρόβλεψη δραστηριοτήτων επιχειρησιακών διεργασιών με πιθανοτικά μοντέλα μάθησης.

Πρασίδης Ιωάννης
A.M. 71346373

Επιβλέπων Καθηγητής: Μπουσδέκης Αλέξανδρος

Συν-επιβλέπων καθηγητής: Μιαούλης Γεώργιος

ΑΘΗΝΑ-ΑΙΓΑΛΕΩ, 07/2021



UNIVERSITY OF WEST ATTICA
FACULTY OF ENGINEERING
DEPARTMENT OF INFORMATICS & COMPUTER
ENGINEERING

Diploma Thesis

Activity prediction of business processes with probabilistic learning models.

Prasidis Ioannis
Registration Number: 71346373

Supervisor: Bousdekis Alexandros

Co-supervisor: Miaoulis Georgios

ATHENS-EGALEO, 07/2021

ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

Πρόβλεψη δραστηριοτήτων επιχειρησιακών διεργασιών με πιθανοτικά μοντέλα μάθησης.

**Πρασίδης Ιωάννης
Α.Μ. 71346373**

Επιβλέπων Καθηγητής: Μπουσδέκης Αλέξανδρος

Συν-επιβλέπων καθηγητής: Μιαούλης Γεώργιος

Εξεταστική Επιτροπή:

Μιαούλης Γεώργιος: Καθηγητής ΠΑΔΑ

Μαστοροκόστας Πάρις: Καθηγητής ΠΑΔΑ

Βουλόδημος Αθανάσιος: Επίκουρος Καθηγητής ΠΑΔΑ

Ημερομηνία εξέτασης 21/07/2021

Δήλωση Συγγραφέα Διπλωματικής Εργασίας

Ο κάτωθι υπογεγραμμένος Πρασίδης Ιωάννης του Κωνσταντίνου, με αριθμό μητρώου 71346373 φοιτητής του Πανεπιστημίου Δυτικής Αττικής της Σχολής Μηχανικών του Τμήματος Μηχανικών Πληροφορικής και Υπολογιστών, δηλώνω υπεύθυνα ότι:

«Είμαι συγγραφέας αυτής της 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



Περίληψη

Οι διαδικασίες αποτελούν το πιο σημαντικό στοιχείο μιας επιχείρησης, καθώς η επιτυχής εκτέλεση τους συμβάλλει στην ολοκλήρωση των οργανωτικών της στόχων. Προκειμένου όμως να είναι σε θέση να πετυχαίνει τους στόχους της, πρέπει να γίνεται ανάλυση των διαδικασιών αυτών για τυχόν αποκλίσεις αλλά και να δημιουργηθεί ένα μοντέλο το οποίο θα περιγράφει αναλυτικά το σύνολο των δραστηριοτήτων που εκτελούνται, αλλά και των σχέσεων μεταξύ των. Μέσω των παραπάνω τεχνικών, γίνεται ευκολότερη η κατανόηση των διαδικασιών που πραγματοποιούνται σε μια επιχείρηση, των προβλημάτων αλλά και των συνεισφορών που μπορεί να προσφέρει μια αλλαγή.

Δεν είναι λίγες οι φορές που, κάποιο μοντέλο δεν «ταυτίζεται» με τις υπάρχουσες διαδικασίες της επιχείρησης, καθώς είτε υπάρχουν σχέσεις μεταξύ δραστηριοτήτων που φαινομενικά δεν μπορούν να πραγματοποιηθούν είτε δεν υπάρχουν στο σύστημα της επιχείρησης. Και όλα αυτά γιατί από μόνη της μια διαδικασία μπορεί να γίνει αρκετά πολύπλοκη από τη φύση της, όταν προσπαθήσουμε να την εφαρμόσουμε σε κάποιο σύστημα. Επομένως, χρήζει αντιμετώπισης η αβεβαιότητα που συναντάμε στις επιχειρησιακές διαδικασίες, μέσω κάποιου πιθανοτικού μοντέλου μάθησης.

Ειδικότερα, η διπλωματική εργασία αφορά την εξόρυξη επιχειρησιακών διεργασιών και την πρόβλεψη δραστηριοτήτων με τη χρήση πιθανοτικών μοντέλων μάθησης.

Πιο συγκεκριμένα, το πρώτο βήμα της προσέγγισης είναι η εξαγωγή των μοντέλων διεργασιών (process models) και των δικτύων Petri (Petri nets) από αρχεία καταγραφής συμβάντων (event logs) με τη χρήση των αλγορίθμων εξόρυξης διεργασιών. Έπειτα, γίνεται ανάλυση των παραπάνω με σκοπό την περεταίρω μελέτη και προβολή επιμέρους λεπτομερών σχετικά με τα αρχεία καταγραφής συμβάντων (event logs) αλλά και πιο συγκεκριμένα των διεργασιών που εμπεριέχουν. Στη συνέχεια, τα events logs και το process model εισέρχονται σε ένα Μπεϋζιανό δίκτυο (Bayesian Network), προκειμένου να δημιουργηθεί ένα πιθανοτικό μοντέλο, το οποίο μπορεί να «μαθαίνει» κάθε φορά που θα εισάγονται νέες εγγραφές.

Με αυτόν τον τρόπο, η προσέγγιση της διπλωματικής επιτρέπει τη διαχείριση της αβεβαιότητας, που αφορά τη μετάβαση στην επόμενη δραστηριότητα, με τον υπολογισμό των αντίστοιχων πιθανοτήτων κατά τον νόμο του Μπέυζ (Bayes).

Η υλοποίηση γίνεται με τη χρήση της γλώσσας προγραμματισμού Python.

Λέξεις κλειδιά: εξόρυξη, επιχειρησιακές διεργασίες, πρόβλεψη δραστηριοτήτων, πιθανοτικά μοντέλα μάθησης, μοντέλο διεργασιών, αρχεία καταγραφής συμβάντων, αλγόριθμοι εξόρυξης διεργασιών, μπεϋζιανό δίκτυο, πιθανοτικό μοντέλο, νόμος Μπέυζ.

Abstract

Processes are the most important element of a business, as their successful execution contributes to the completion of its organizational goals. However, in order to be able to achieve its goals, these processes must be analyzed for any deviations and a model must be created that will describe in detail all the activities performed, but also the relationships between them. Through the above techniques, it becomes easier to understand the processes that take place in a business, the problems and the contributions that a change can offer.

It is not uncommon for a model not to "identify" with existing business processes, as there are either relationships between activities that seemingly cannot be accomplished or do not exist in the business system. And all this because a process can become quite complex in nature, when we try to apply it to a system. Therefore, the uncertainty we encounter in business processes needs to be addressed, through some probabilistic learning model. In particular, this thesis discusses the business process mining and the activity prediction with probabilistic learning models.

More specifically, the first step in the approach is to export process models and Petri nets from the event log using process mining algorithms. Then, the above is analyzed in order to further study and view individual details about the event logs and more specifically the processes involved. Afterwards, the event log and process model are entered into a Bayesian Network in order to create a probabilistic model that can "learn" each time new entries are entered.

In this way, this thesis approach allows the uncertainty, regarding the transition to the next activity, to be managed by calculating the corresponding probabilities under Bayes law.

The implementation is done using the Python programming language

Keywords: mining, business process, activity prediction, probabilistic learning models, process model, event log, process mining algorithms, Bayesian Network, probabilistic model, Bayes law.

Πίνακας περιεχομένων

Περίληψη	1
Abstract	3
Πίνακας περιεχομένων	5
Κατάλογος εικόνων	8
Κατάλογος πινάκων	10
Πίνακας ακρωνύμιων	11
1. Εισαγωγή	12
1.1 Γενικά	12
1.2 Αντικείμενο και Στόχοι της Διπλωματικής Εργασίας	12
1.3 Διάρθρωση της Διπλωματικής Εργασίας	12
2. Βιβλιογραφική Επισκόπηση	15
2.1 Διαχείριση Επιχειρησιακών Διαδικασιών (Business Process Management)	15
2.1.1 Ιστορική αναδρομή.	16
2.1.2 Βασικοί ορισμοί.	16
2.1.3 Παραδοσιακή δομή λειτουργίας επιχειρήσεων.....	18
2.1.4 Στροφή των επιχειρήσεων σε BPM.	21
2.1.5 Βήματα Διαχείρισης Επιχειρησιακών Διαδικασιών (BPM lifecycle).....	22
2.1.6 Πότε πρέπει μια επιχείρηση να ασχοληθεί με το BPM.....	24
2.2 Εξόρυξη Διαδικασιών (Process Mining)	27
2.2.1 Εισαγωγή στην εξόρυξη διαδικασιών, τύποι, οπτική γωνία και προβλήματα. 27	
2.2.1.1 Εισαγωγή	27
2.2.1.2 Τύποι εξόρυξης διεργασιών.....	28
2.2.1.3 Οπτικές γωνίες εξόρυξης διαδικασιών.	30
2.2.1.4 Συνήθη προβλήματα στην εξόρυξη διαδικασιών	30
2.2.1.5 Δίκτυα Petri (Petri nets).....	31
2.2.1.6 Μοντελοποίηση επιχειρησιακών διαδικασιών (BPMN)	33
2.2.2 Αρχεία καταγραφής γεγονότων, μοντέλα διεργασιών και οι σχέσεις μεταξύ τους (Play-Out, Play-In , Replay).	35
2.2.2.1 Αρχεία καταγραφής γεγονότων – event logs.....	35
2.2.2.2 Μοντέλα διεργασιών – process models.....	38

2.2.2.3	Σχέσεις μεταξύ αρχείου καταγραφής συμβάντων και μοντέλου διεργασιών.....	39
2.2.3	Ανακάλυψη.....	41
2.2.3.1	Περιπτώσεις εφαρμογής της τεχνικής ανακάλυψης διαδικασιών.....	41
2.2.3.2	Αλγόριθμοι ανακάλυψης μοντέλων διαδικασιών.....	42
2.2.4	Κριτήρια αξιολόγησης ενός μοντέλου διαδικασιών.....	43
2.2.5	Έλεγχος συμμόρφωσης.....	45
2.2.5.1	Περιπτώσεις εφαρμογής ελέγχου συμμόρφωσης.....	46
2.2.5.2	Τεχνικές ελέγχου συμμόρφωσης.....	47
2.3	Μπεϋζιανά Δίκτυα (Bayesian Networks).....	50
2.3.1	Εισαγωγή.....	50
2.3.2	Θεωρία πιθανοτήτων.....	50
2.3.3	Θεωρητικό υπόβαθρο Bayesian Networks.....	52
2.3.4	Θεώρημα Bayes.....	54
2.3.5	Συμπερασμός (Inference).....	55
3.	Προτεινόμενη προσέγγιση.....	57
3.1	Μέρος Α.....	61
3.1.1	Διαδικασία ανακάλυψης επιχειρησιακού μοντέλου.....	61
3.1.2	Έλεγχος συμμόρφωσης (Conformance checking).....	61
3.1.3	Αξιολόγηση επιχειρησιακών μοντέλων (Evaluation).....	62
3.1.4	Περαιτέρω αναγνώσεις πληροφοριών αρχείων καταγραφής συμβάντων.....	63
3.2	Μέρος Β.....	63
3.2.1	Εξαγωγή άκυκλου κατευθυνόμενου γραφήματος.....	63
3.2.2	Αφαίρεση κύκλων.....	64
3.2.3	Δημιουργία Μπεϋζιανού δικτύου.....	65
3.2.4	Υποθέσεις και πρόβλεψη δραστηριότητας.....	66
4.	Μελέτη Περίπτωσης: Αίτημα αποζημίωσης πτήσης σε αεροπορική εταιρία.....	67
4.1	Περί event log.....	68
4.2	Ανακάλυψη επιχειρησιακού μοντέλου.....	70
4.3	Έλεγχος συμμόρφωσης στην πράξη.....	73
4.4	Αξιολόγηση των επιχειρησιακών μοντέλων.....	78
4.5	Λοιπές πληροφορίες για το αρχείο καταγραφής συμβάντων (statistics, decision mining).....	81
4.6	Προετοιμασία για τη δημιουργία Μπεϋζιανού μοντέλου.....	83

4.7	Αναζήτηση κύκλων	84
4.8	Αφαίρεση κύκλων	85
4.9	Κατασκευή Μπεϋζιανού μοντέλου	86
4.10	Ανάλυση και αποτελέσματα	90
5.	Συμπεράσματα και Μελλοντική Εργασία	94
	Βιβλιογραφικές Αναφορές	97
	Ελληνικές βιβλιογραφικές αναφορές	97
	Ξενόγλωσσες βιβλιογραφικές αναφορές	97

Κατάλογος εικόνων

Εικόνα 1. Παράδειγμα οργανογράμματος.....	19
Εικόνα 2. Απεικόνιση των διεργασιών.	21
Εικόνα 3. Παράδειγμα οριζόντιας οργανωτικής δομής.	22
Εικόνα 4. BPM lifecycle.	23
Εικόνα 5. Η εξόρυξη διεργασιών αποτελεί τον σύνδεσμο μεταξύ της επιστήμης δεδομένων και της επιστήμης διεργασιών (W. Van der Aalst, 2016).	27
Εικόνα 6. Τα «περιεχόμενα» της εξόρυξης διεργασιών.	28
Εικόνα 7. Οι τρεις τύποι εξόρυξης διαδικασιών σε μορφή εισόδου/εξόδου.	29
Εικόνα 8. Παράδειγμα Petri net αποτελούμενο από 8 μέρη και 9 μεταβάσεις. [9]	32
Εικόνα 9. Εκ νέου παράδειγμα Petri net.	33
Εικόνα 10. Παράδειγμα διαγράμματος BPMN. [9]	34
Εικόνα 11. Σύνολο συμβόλων που χρησιμοποιεί το BPMN. [9]	35
Εικόνα 12. Παράδειγμα μέρους αρχείου καταγραφής γεγονότων.	36
Εικόνα 13. Παράδειγμα μέρους XES αρχείου.	37
Εικόνα 14. Play-In, Play-Out και Replay. [9]	40
Εικόνα 15. Στα αριστερά απεικονίζεται ένα αρχείο καταγραφής γεγονότων και στα δεξιά απεικονίζονται δύο ενδεικτικά μοντέλα διαδικασιών που περιγράφουν τη συμπεριφορά που παρατηρήθηκε στο αρχείο. [24]	41
Εικόνα 16. Διαστάσεις ποιότητας μοντέλου διαδικασιών. [21].....	44
Εικόνα 17. Παράδειγμα αποτυπώματος.	47
Εικόνα 18. Παραδείγματα τεχνικής alignments.....	48
Εικόνα 19. Παράδειγμα δομής μπεϋζιανού δικτύου. [28]	53
Εικόνα 20. Παράδειγμα BN για συμπερασμό (chair = c, sport = s, worker = w, back = b, ache = a). [32]	55
Εικόνα 21. Αλγόριθμος για εξαγωγή μοντέλου διεργασιών και περισσότερα σε εξόρυξη διεργασιών.....	58
Εικόνα 22. Αλγόριθμος για μετατροπή αρχείου καταγραφής συμβάντων σε Bayesian Network.....	60
Εικόνα 23. Τα τέσσερα ανταγωνιστικά ποιοτικά κριτήρια. [9]	63
Εικόνα 24. Αποσύνθεση κύκλων μεγέθους 1, 2, ν, αντίστοιχα. [12].....	65
Εικόνα 25. Μοντέλο διεργασιών παραγόμενο από τον Alpha Miner (a = register request, b = examine thoroughly, c = examine casually, d = check ticket, e = decide, f = reinstantiate request, g = pay compensation, h = reject request).	71
Εικόνα 26. Μοντέλο διεργασιών παραγόμενο από τον Inductive Miner (a = register request, b = examine thoroughly, c = examine casually, d = check ticket, e = decide, f = reinstantiate request, g = pay compensation, h = reject request).	71
Εικόνα 27. Μοντέλο διεργασιών παραγόμενο από τον Heuristic Miner (a = register request, b = examine thoroughly, c = examine casually, d = check ticket, e = decide, f = reinstantiate request, g = pay compensation, h = reject request).	72

Εικόνα 28. Μοντέλο διεργασιών Directly-Follows Graph (a = register request, b = examine thoroughly, c = examine casually, d = check ticket, e = decide, f = reinstantiate request, g = pay compensation, h = reject request).	73
Εικόνα 29. Replaying the case a-c-d-e-h. (p = produced tokens, c = consumed tokens, m = missing tokens, r = remaining tokens).	74
Εικόνα 30. Alignments μοντέλο παραδείγματος.	76
Εικόνα 31. Κατανομή χρονικής διάρκειας μιας περίπτωσης.	81
Εικόνα 32. Κατανομή των δραστηριοτήτων ανά την πάροδο του χρόνου.	81
Εικόνα 33. Decision points on Inductive miner process model.	82
Εικόνα 34. Decision mining results.	83
Εικόνα 35. BPMN ενός μέρους του μοντέλου εξαγόμενο από τον inductive miner.	84
Εικόνα 36. Εκ νέου Directly follows graph.	85
Εικόνα 37. Directed Acyclic Graph (DAG).	86
Σε αυτό το σημείο το μπεϋζιανό δίκτυο έχει ολοκληρωθεί, αφού έχουμε και άκυκλο κατευθυνόμενο γράφημα (εικόνα 38) και τους πίνακες δεσμευμένων πιθανοτήτων (CBT) (πίνακες 17 έως 23).	89
Εικόνα 39. Confusion matrix.	91

Κατάλογος πινάκων

Πίνακας 1. Βασικές πληροφορίες αρχείου καταγραφής συμβάντων.....	68
Πίνακας 2. Σύνολο των περιπτώσεων (a = register request, b = examine thoroughly, c = examine casually, d = check ticket, e = decide, f = reinitiate request, g = pay compensation, h = reject request).....	69
Πίνακας 3. Μέρος του αρχείου καταγραφής σε μορφή xes.....	70
Πίνακας 4. Αποτελέσματα token-based συμμόρφωσης.....	75
Πίνακας 5. Alignments παράδειγμα 1.....	76
Πίνακας 6. Alignments παράδειγμα 2.....	77
Πίνακας 7. Αποτελέσματα alignments συμμόρφωσης.....	78
Πίνακας 8. Αξιολόγηση μοντέλου Alpha Miner.....	78
Πίνακας 9. Αξιολόγηση μοντέλου Inductive Miner.	79
Πίνακας 10. Αξιολόγηση μοντέλου Heuristic Miner.	79
Πίνακας 11. Αξιολόγηση μοντέλου Directly-Follows Graph.	80
Πίνακας 12. Σύγκριση των μοντέλων.	80
Πίνακας 13. Χρονικές λεπτομέρειες του αρχείου καταγραφής συμβάντων.	82
Πίνακας 14. Ακολουθία δραστηριοτήτων που ανήκουν σε κύκλο.	85
Πίνακας 15. Βασικές πληροφορίες φιλτραρισμένου αρχείου καταγραφής συμβάντων.	87
Πίνακας 16. Πίνακας 2. Σύνολο των περιπτώσεων φιλτραρισμένου αρχείου καταγραφής συμβάντων.....	87
Πίνακας 17. CBT of a.	88
Πίνακας 18. CBT of b.	88
Πίνακας 19. CBT of c.	88
Πίνακας 20. CBT of d.	88
Πίνακας 21. CBT of d'.	89
Πίνακας 22. CBT of e.	89
Πίνακας 23. CBT of output activity (g/h).	89
Πίνακας 24. Υπολογισμός πιθανοτήτων εμφάνισης μιας δραστηριότητας ενώ γνωρίζουμε ότι έχει εκτελεστεί μια άλλη δραστηριότητα.	90
Πίνακας 25. Ανάλυση πίνακα σύγχυσης (confusion matrix).....	91

Πίνακας ακρωνύμιων

BPM	Business Process Management
TQM	Total Quality Management
BPR	Business Process Reengineering
ERP	Enterprise Resource Planning
CRM	Customer Relation Management
6σ	Six Sigma
XES	eXtensible Event Stream
MXML	Mining eXtensible Markup Language
CSV	Comma-Separated Values
IEEE	Institute of Electrical and Electronics Engineers
BPMN	Business Process Model and Notation
UML	Unified Modeling Language
BPMI	Business Process Management Initiative
OMG	Object Management Group
BPD	Business Process Diagram
BN	Bayesian network
DAG	Directed Acyclic Graph
CPT	Conditional Probability Table

1. Εισαγωγή

1.1 Γενικά

Η εξόρυξη επιχειρησιακών διεργασιών υπόκειται σε ένα πρόσφατο αλλά και επεκτεινόμενο τομέα έρευνας. Κύριος στόχος είναι η ανάλυση των επιχειρησιακών διαδικασιών, μέσω των οποίων μπορούμε να κατανοήσουμε πως εκτελούνται οι διαδικασίες σε μια επιχείρηση και να προσπαθήσουμε να τις βελτιώσουμε.

Συχνό φαινόμενο είναι η ύπαρξη της αβεβαιότητας. Αβεβαιότητας μεταξύ των διαδικασιών, περιπτώσεις που δεν μπορούμε να γνωρίζουμε εάν έχει πραγματοποιηθεί μια διεργασία και θέλουμε να μάθουμε την πιθανότητα της επιχειρησιακής διαδικασίας να τερματιστεί ή να φτάσει σε μια επόμενη διεργασία.

Εδώ λοιπόν βρίσκεται και η καινοτομία της διπλωματικής. Η ιδιότητα να συνδυάζουμε την εξόρυξη επιχειρησιακών διεργασιών με τα Μπεϋζιανά δίκτυα με σκοπό την αντιμετώπιση της αβεβαιότητας και την μετάβαση στην επόμενη δραστηριότητα, μιας και ένα μπεϋζιανό δίκτυο μπορεί και λαμβάνει υπόψη όλες τις πιθανότητες, παρουσίας ή και μη, μιας εργασίας σε μια επιχειρησιακή διεργασία.

1.2 Αντικείμενο και Στόχοι της Διπλωματικής Εργασίας

Αντικείμενο της παρούσας διπλωματικής εργασίας είναι η ανάπτυξη και εφαρμογή μιας προσέγγισης, η οποία μέσω της εξόρυξης διαδικασιών θα ανακαλύπτει και αναλύει τις διαδικασίες μιας επιχείρησης και έπειτα θα δημιουργείται ένα μπεϋζιανό δίκτυο που θα περιγράφει τις παραπάνω διαδικασίες, με απώτερο στόχο την πρόβλεψη εμφάνισης κάποιας επιχειρησιακής δραστηριότητας.

1.3 Διάρθρωση της Διπλωματικής Εργασίας

Η παρούσα διπλωματική εργασία διαρθρώνεται ως εξής:

- **Κεφάλαιο 2: Βιβλιογραφική επισκόπηση:** Στο παρόν κεφάλαιο αναλύεται το θεωρητικό υπόβαθρο της διπλωματικής εργασίας. Χωρίζεται σε 3 υποκεφάλαια,

στα οποία ξεχωριστά αναλύονται οι τρεις τομείς που στηρίζεται η διπλωματική εργασία.

○ Κεφάλαιο 2.1: Διαχείριση επιχειρησιακών διαδικασιών (BPM):

Στην ενότητα 2.1, αναλύονται οι όροι «διαδικασία», «επιχειρησιακή διαδικασία» και «διαχείριση επιχειρησιακών διαδικασιών». Παρουσιάζεται η δομή λειτουργίας των επιχειρήσεων και η ανάγκη για αλλαγή της παραδοσιακή τους οργάνωσης, σε οργάνωση με κέντρο τις διαδικασίες. Ακολουθούν τα βήματα για μια σωστή διαχείριση επιχειρησιακών διαδικασιών, πότε πρέπει μια επιχείρηση να εφαρμόσει τη διαχείριση αυτή καθώς και τα οφέλη που προσφέρει.

○ Κεφάλαιο 2.2: Εξόρυξη διαδικασιών (PM):

Στην ενότητα 2.2, ορίζεται η εξόρυξη διεργασιών. Γίνεται μια προσπάθεια για την αποσαφήνιση του όρου αυτού και θα δείχνονται οι δυνατότητές του. Συνεχίζοντας, γίνεται λόγος για τα αρχεία καταγραφής συμβάντων (event logs), δίκτυα Petri (Petri nets) και τα μοντέλα διαδικασιών (process models). Αναδεικνύονται οι κύριοι αλγόριθμοι εξαγωγής τέτοιων μοντέλων και γίνεται αναφορά στον τρόπο εξέτασής των ως προς την ποιότητά τους. Τέλος αναλύεται μια ακόμα τεχνική ελέγχου συμμόρφωσης που ελέγχει τέτοια μοντέλα αν συμμορφώνονται με το αρχείο καταγραφής συμβάντων, γνωστή και ως conformance checking.

○ Κεφάλαιο 2.3: Μπεϋζιανά δίκτυα (BN):

Στην ενότητα 2.3, αναφέρονται τα Μπεϋζιανά δίκτυα. Γίνεται μια εισαγωγή γύρω από τη χρησιμότητά τους, δίνεται ο ορισμός τους καθώς και ένα παράδειγμα, με σκοπό την ευκολότερη κατανόησή τους. Επίσης, αναλύονται σε έκταση όλες οι μαθηματικές αρχές πίσω από τα Μπεϋζιανά δίκτυα. Δεν παραλείπεται η αναφορά σε μπεϋζιανούς ταξινομητές καθώς και στη γνωστή μέθοδο του συμπερασμού. Τέλος, βλέπουμε πως μπορούμε να τα «συνδυάσουμε» με τις επιχειρησιακές διεργασίες και την εξόρυξη διεργασιών. Πως μπορούμε να μετατρέψουμε ένα σύνολο διεργασιών σε ένα Μπεϋζιανό δίκτυο και να προχωρήσουμε από εκεί με στόχο κάποιο από τα προτερήματα που μας προσφέρει, είτε αυτό πρόκειται για κάποια πρόβλεψη, χειρισμό δεδομένων με ελλειπίς καταχωρήσεις ή κάτι άλλο.

- Κεφάλαιο 3: Προτεινόμενη προσέγγιση: Στο παρόν κεφάλαιο περιγράφεται η προτεινόμενη προσέγγιση που ακολουθήθηκε. Προτείνεται ένας αλγόριθμος χειρισμού της διπλωματικής εργασίας και των μερών που αποτελείται. Χωρίζεται σε δύο μέρη, όπου στο πρώτο εφαρμόζονται τεχνικές εξόρυξης διαδικασιών με στόχο την ανακάλυψη και ανάλυση των επιχειρησιακών διεργασιών και στο δεύτερο και κυριότερο μέρος, ορίζεται η μετατροπή κάποιου επιχειρησιακού μοντέλου, που έχουμε πάρει από το πρώτο μέρος της προσέγγισης, σε ένα μπεϋζιανό δίκτυο, με απώτερο στόχο την πρόβλεψη εμφάνισης κάποιας δραστηριότητας.
- Κεφάλαιο 4: Μελέτη περίπτωσης: Στο παρόν κεφάλαιο, εφαρμόζεται η προτεινόμενη προσέγγιση σε ένα αληθινό σενάριο και παρουσιάζονται αναλυτικά τα αποτελέσματα.
- Κεφάλαιο 5: Συμπεράσματα και μελλοντική εργασία.
- Βιβλιογραφικές αναφορές.

2. Βιβλιογραφική Επισκόπηση

Στο παρόν κεφάλαιο αναλύεται το θεωρητικό υπόβαθρο της διπλωματικής εργασίας.

Αρχικά, στην ενότητα 2.1, θα αναλυθούν οι όροι «διαδικασία», «επιχειρησιακή διαδικασία» και «διαχείριση επιχειρησιακών διαδικασιών». Παρουσιάζεται η δομή λειτουργίας των επιχειρήσεων και η ανάγκη για αλλαγή της παραδοσιακή τους οργάνωσης, σε οργάνωση με κέντρο τις διαδικασίες. Θα ακολουθήσουν τα βήματα για μια σωστή διαχείριση επιχειρησιακών διαδικασιών, τότε πρέπει μια επιχείρηση να εφαρμόσει τη διαχείριση αυτή καθώς και τα οφέλη που προσφέρει.

Στην ενότητα 2.2, ορίζουμε την εξόρυξη διεργασιών. Θα γίνει μια προσπάθεια για την αποσαφήνιση του όρου αυτού και θα δειχθούν οι δυνατότητές του. Συνεχίζοντας, θα μιλήσουμε για τα αρχεία καταγραφής συμβάντων (event logs), δίκτυα Petri (Petri nets) και τα μοντέλα διαδικασιών (process models). Θα αναδειχθούν οι κύριοι αλγόριθμοι εξαγωγής τέτοιων μοντέλων και θα γίνει αναφορά στον τρόπο εξέτασής των ως προς την ποιότητά τους. Τέλος θα γίνει λόγος μιας ακόμα τεχνικής ελέγχου συμμόρφωσης που ελέγχει τέτοια μοντέλα αν συμμορφώνονται με το αρχείο καταγραφής συμβάντων, γνωστή και ως conformance checking.

Στην ενότητα 2.3, αναφέρονται τα Μπεϋζιανά δίκτυα. Γίνεται μια εισαγωγή γύρω από τη χρησιμότητά τους, δίνεται ο ορισμός τους καθώς και ένα παράδειγμα, με σκοπό την ευκολότερη κατανόησή τους. Επίσης, αναλύονται σε έκταση όλες οι μαθηματικές αρχές πίσω από τα Μπεϋζιανά δίκτυα. Δεν παραλείπεται η αναφορά και στη γνωστή μέθοδο του συμπερασμού. Τέλος, θα δούμε πως μπορούμε να τα «συνδυάσουμε» με τις επιχειρησιακές διεργασίες και την εξόρυξη διεργασιών. Πως μπορούμε να μετατρέψουμε ένα σύνολο διεργασιών σε ένα Μπεϋζιανό δίκτυο και να προχωρήσουμε από εκεί με στόχο κάποιο από τα προτερήματα που μας προσφέρει, είτε αυτό πρόκειται για κάποια πρόβλεψη, χειρισμό δεδομένων με ελλειπίς καταχωρήσεις ή κάτι άλλο.

2.1 Διαχείριση Επιχειρησιακών Διαδικασιών (Business Process Management)

2.1.1 Ιστορική αναδρομή.

Η πορεία προς τη διαχείριση επιχειρησιακών διαδικασιών (BPM) ανά τα χρόνια έχει περάσει αρκετές δυσκολίες. Έχουν γίνει πολλές προσπάθειες, επιτυχημένες και μη, για την επίτευξη οργανωτικής αποτελεσματικότητας με βάση της διαδικασίες που έχουν διαμορφώσει σήμερα τη διαχείριση επιχειρησιακών διαδικασιών. [6]

Τη δεκαετία του 1980, το επίκεντρο ήταν η διαχείριση της συνολικής ποιότητας (TQM). Ερχόμενοι στη δεκαετία του 1990, στραφήκαμε στον επανασχεδιασμό των επιχειρησιακών διεργασιών (BPR), μια ιδέα των Hammer και Champy. Ακόλουθος του BPR στα τέλη του 1990 ήταν τα συστήματα ενδοεπιχειρησιακού σχεδιασμού (ERP) τα οποία εστιάζοντας στην οργάνωση έγιναν το επόμενο μεγάλο πράγμα. Υποτίθεται ότι πρόσφεραν βελτιωμένους τρόπους λειτουργίας των επιχειρήσεων και πωλούνταν ως «λύση στα προβλήματά σας», πράγμα που δεν έγινε. Κατά τις αρχές του 2000, βγήκαν στην αγορά πολλά συστήματα που βασίζονταν στη διαχείριση των πελατειακών σχέσεων (CRM), δίνοντας έμφαση στην άποψη των πελατών. Ακόμα όμως δεν ήταν αρκετό. Έπειτα από τα CRM εμφανίστηκε το Six Sigma (6σ). [6]

2.1.2 Βασικοί ορισμοί.

Πριν προχωρήσουμε στο να κατονομάσουμε την διαχείριση επιχειρησιακών διαδικασιών, θεωρείται σκόπιμη η αποσαφήνιση των βασικών όρων που περιλαμβάνει.

Πολλοί έχουν προσπαθήσει να δώσουν από έναν ορισμό για την έννοια της διαδικασίας, δημιουργώντας έτσι ένα μεγάλο φάσμα περιπτώσεων ορισμών, καλύπτοντας πολλές πλευρές του ανάλογα με την οπτική γωνία του καθενός. Ορισμένοι από τους ορισμούς είναι οι παρακάτω:

- “Διαδικασία είναι μια σειρά ενεργειών που έχουν ληφθεί για την επίτευξη ενός συγκεκριμένου σκοπού.” [1]
- “Διαδικασία είναι ένα φυσικό φαινόμενο που χαρακτηρίζεται από βαθμιαίες αλλαγές που οδηγούν σε ένα συγκεκριμένο αποτέλεσμα.” [2]
- “Διαδικασία είναι μια σειρά ή ένα σύνολο δραστηριοτήτων που αλληλοεπιδρούν για να παράγουν ένα αποτέλεσμα. Μπορεί να συμβεί μόνο μία φορά ή να είναι επαναλαμβανόμενο ή περιοδικό.” [3]

Με βάση τα παραπάνω παρατηρούμε τα χαρακτηριστικά της διαδικασίας, που ουσιαστικά πρόκειται για ένα ορισμένο περιβάλλον, δηλαδή έχει αρχή και τέλος και δραστηριότητες που αλληλοεπιδρούν μεταξύ τους χρονικά δίνοντας κάποια έξοδο, ένα αποτέλεσμα.

Συνεχίζουμε με τον όρο επιχειρησιακή διαδικασία. Ο (Weske, 2019), έκανε προσπάθειες για τον κατατοπισμό του όρου, μια εκ των οποίων βλέπουμε παρακάτω.

Ένας πρώτος ορισμός που έδωσε λέει ότι, “μια επιχειρησιακή διαδικασία αποτελείται από ένα σύνολο δραστηριοτήτων που εκτελούνται συντονισμένα σε ένα οργανωτικό και τεχνικό περιβάλλον. Αυτές οι δραστηριότητες συνειδητοποιούν από κοινού έναν επιχειρηματικό στόχο. Κάθε επιχειρηματική διαδικασία εφαρμόζεται από έναν μεμονωμένο οργανισμό, αλλά μπορεί να αλληλοεπιδρά με τις επιχειρηματικές διαδικασίες που εκτελούνται από άλλους οργανισμούς.” [4]

Οι (Rosing et al., 2014) ορίζουν την επιχειρησιακή διαδικασία ως “είναι η ίδια με μια διαδικασία ωστόσο με μια σημαντική διαφορά, την έμφαση στη λέξη επιχείρηση. Μια επιχειρηματική διαδικασία δηλαδή, είναι μια συλλογή εργασιών και δραστηριοτήτων (επιχειρησιακές λειτουργίες και ενέργειες) που αποτελούνται από υπαλλήλους, υλικά, μηχανήματα, συστήματα και μεθόδους που είναι δομημένες με τέτοιο τρόπο ώστε να σχεδιάζουν, δημιουργούν και να παρέχουν ένα προϊόν ή μια υπηρεσία στον καταναλωτή”. [5]

Ουσιαστικά, πρόκειται για μια σειρά ενεργειών που εκτελούνται από ένα ή και παραπάνω πρόσωπα, ατομικά ή σε ομάδα, για τη δημιουργία ενός προϊόντος ή κάποιας υπηρεσίας προς κάποιον πελάτη.

Τέλος, όσον αφορά τη διαχείριση επιχειρησιακών διαδικασιών, υπάρχουν πολλοί ορισμοί οι οποίοι διαφέρουν μεταξύ τους. Ας δούμε λοιπόν κάποιους από τους ορισμούς που έχουν δοθεί μέχρι στιγμής και στο τέλος θα γίνει μια προσπάθεια μέσω των ομοιοτήτων τους να γίνει πιο κατανοητός ο όρος, δίνοντας μια πιο σφαιρική άποψη πάνω στο θέμα. Σύμφωνα με τον (Weske, 2019), “η διαχείριση επιχειρηματικών διαδικασιών περιλαμβάνει έννοιες, μεθόδους και τεχνικές για την υποστήριξη του σχεδιασμού, της διαχείρισης, της διαμόρφωσης, της εφαρμογής και της ανάλυσης επιχειρηματικών διαδικασιών”. [4]

Οι (Jeston & Nelis, 2008) ορίζουν την διαχείριση επιχειρηματικών διαδικασιών ως “ την επίτευξη των στόχων ενός οργανισμού μέσω της βελτίωσης, της διαχείρισης και τον έλεγχο των βασικών επιχειρηματικών διαδικασιών”. [6]

Στο γνωστό γλωσσάριο (Gartner, 2021) βρίσκουμε ότι η διαχείριση επιχειρηματικών διαδικασιών “είναι ένας κλάδος που χρησιμοποιεί διάφορες μεθόδους για να ανακαλύψει, να μοντελοποιήσει, να αναλύσει, να μετρήσει, να βελτιώσει και να βελτιστοποιήσει τις επιχειρηματικές διαδικασίες”. [7]

Στην (BPM.com, 2021) θα βρούμε ότι πρόκειται για “ μια πειθαρχία που περιλαμβάνει οποιονδήποτε συνδυασμό μοντελοποίησης, αυτοματισμού, εκτέλεσης, ελέγχου, μέτρησης και βελτιστοποίησης ροών επιχειρηματικής δραστηριότητας, για την υποστήριξη επιχειρηματικών στόχων, συστημάτων, υπαλλήλων, πελατών και συνεργατών εντός και πέρα από τα όρια της επιχείρησης”. [8]

Τι είναι τελικά η διαχείριση επιχειρηματικών διαδικασιών και τι θα κρατήσουμε από τα παραπάνω;

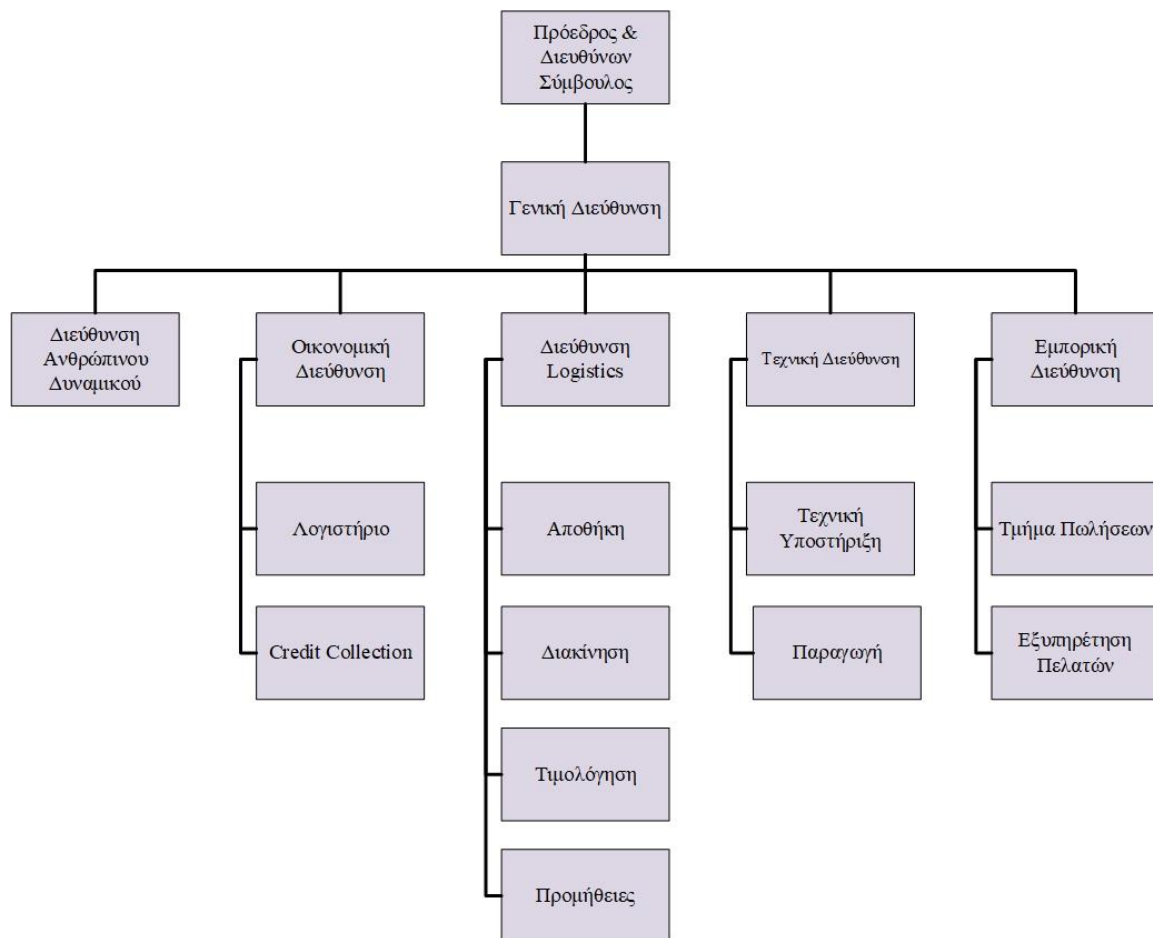
Δεν πρόκειται για κάποιο πρόγραμμα που μπορεί να αγοράσει μια επιχείρηση και να βελτιώσει τις διαδικασίες της από τη μια μέρα στην άλλη, αλλά για μια πειθαρχία, κάτι που το αποκτάς έπειτα από ώρες πρακτικής εξάσκησης. Μέσα από διάφορες τεχνικές, θα πρέπει μια επιχείρηση να καταφέρει να «δει» τις διαδικασίες της από το μηδέν, να φτιάξει ένα μοντέλο που να τις αντιπροσωπεύει, να γίνει σοβαρή ανάλυση πάνω τους με στόχο να τις βελτιώσουν καθώς και να προσπαθήσουν να τις αυτοματοποιήσουν. Οποιαδήποτε ενέργεια γίνει από την επιχείρηση πάνω στις διαδικασίες της, με στόχο την βελτιστοποίηση τους (μείωση χρόνου εκτέλεσης, σφαλμάτων, κόστους κ.α.), είναι μια διαχείριση επιχειρηματικών διαδικασιών.

2.1.3 Παραδοσιακή δομή λειτουργίας επιχειρήσεων.

Κάθε επιχείρηση έχει μια οργανωτική δομή, η οποία ουσιαστικά μας εξηγεί τον τρόπο με τον οποίο μια επιχείρηση διαχειρίζεται το προσωπικό και τις ενέργειες που πρέπει να γίνουν ώστε να καταφέρει να φτάσει τους στόχους της. Ανάλογα με το μέγεθος μιας επιχείρησης, συχνά απουσιάζει όταν αποτελείται από ελάχιστα άτομα, αλλά πάντα

βρίσκεται σε επιχειρήσεις με μεγαλύτερο προσωπικό καθώς περιγράφει τους ρόλους του καθενός και συμβάλλει στην καλύτερη συνεργασία μεταξύ τους.

Ένα τυπικό οργανόγραμμα (εικόνα 1) περιγράφει την οργανωτική δομή μιας επιχείρησης. [16]



Εικόνα 1. Παράδειγμα οργανογράμματος.

Ένα οργανόγραμμα δεν είναι ίδιο για όλες τις επιχειρήσεις και δεν υπάρχει ένα συγκεκριμένο μοντέλο που να είναι ιδανικό για όλες. Υπάρχουν πολλοί παράγοντες που επηρεάζουν τη μορφή του, όπως για παράδειγμα το μέγεθος της επιχείρησης, ο αριθμός των υπηρεσιών που προσφέρει, η γεωγραφική διασπορά, ο εξοπλισμός που χρησιμοποιεί κ.ο.κ. Επομένως κάθε επιχείρηση ανάλογα με τις ανάγκες και τις προδιαγραφές της προσαρμόζει ανάλογα το οργανόγραμμά της. Βέβαια, μολονότι ότι υπάρχουν όλες αυτές οι διαφορές και κάθε οργανόγραμμα είναι θεωρητικά διαφορετικό μεταξύ του, υπάρχουν κάποιες αρχές που βασίζονται όλες οι επιχειρήσεις και παρατηρούνται στην πλειοψηφία

των οργανογραμμμάτων. Ο τύπος της οργανωτικής δομής μιας επιχείρησης μας επιτρέπει να κατανοήσουμε σημαντικά θέματα σχετικά με τον τρόπο λειτουργίας και την οργάνωσή της. [16]

Πριν προχωρήσουμε να δούμε το παραδοσιακό μοντέλο οργανογράμματος που συνηθίζεται να χρησιμοποιούν οι επιχειρήσεις (αν και κάθε επιχείρηση χρησιμοποιεί το δικό της οργανόγραμμα, φτιαγμένο στα μέτρα της, οι τύποι των οργανογραμμμάτων πάνω κάτω είναι συγκεκριμένοι), είναι αξιοσημείωτο να αναφερθούν οι τέσσερις βασικές παράμετροι που διαμορφώνουν σημαντικά μια οργανωτική δομή:

1. Η πρώτη παράμετρος είναι η κατηγοριοποίηση του εργατικού δυναμικού, ο προσδιορισμός των εργασιών στο ανθρώπινο δυναμικό.
2. Η δεύτερη παράμετρος είναι ο καθορισμός των τμημάτων λειτουργίας μιας επιχείρησης.
3. Η τρίτη παράμετρος είναι το εύρος του διοικητικού ελέγχου, δηλαδή ο μέγιστος αριθμός διοικούμενων ατόμων τα οποία μπορεί να κατευθύνει και να επιβλέψει ένας υπάλληλος.
4. Η τέταρτη παράμετρος είναι ο τρόπος διαμοιρασμού δικαιωμάτων λήψης αποφάσεων εντός του οργανισμού.

Έχοντας υπόψη τις τέσσερις παραπάνω παραμέτρους, το πλήθος των τύπων οργανωτικών δομών είναι μεγάλο. Ξεκινώντας με την πρώτη παράμετρο, βρίσκουμε οργανογράμματα παραδοσιακών και γραφειοκρατικών οργανισμών με συγκεκριμένες διεργασίες και οργανογράμματα με ανθρώπινο δυναμικό που πραγματοποιούν εργασίες ανεξαρτήτου τύπου. Στη συνέχεια, με τη δεύτερη παράμετρο ξεχωρίζει η διαμόρφωση τμημάτων ανά λειτουργία σε μια επιχείρηση, για παράδειγμα το σύνολο των οικονομικών εργασιών σε μια επιχείρηση ή οργανισμό ομαδοποιούνται σε ένα νέο τμήμα με τίτλο 'λογιστήριο'. Ακολουθεί η τρίτη παράμετρος που αλληλοεπιδρά με τη δεύτερη και ουσιαστικά πρόκειται για το μέγεθος της ομαδοποίησης που θα γίνει, που εξαρτάται από το πλήθος των ανθρώπων που μπορεί να κατευθύνει και να επιβλέψει ένας προϊστάμενος. Τέλος, η τέταρτη παράμετρος μένει στην κρίση της διοίκησης της επιχείρησης, αν και είναι σύνηθες το φαινόμενο να παραμένει μόνο στα χέρια των πιο ψηλά ιεραρχικών διοικητικών στελεχών. [16]

Η παραδοσιακή δομή λειτουργίας εμφανίζει υψηλή εξειδίκευση καθηκόντων, υπάρχει οργάνωση των τμημάτων με βάση τη λειτουργία, χαμηλός παρατηρείται να είναι ο αριθμός των διοικούμενων ατόμων ανά προϊστάμενο και η λήψη αποφάσεων γίνεται συνήθως από περιορισμένο αριθμό στελεχών που βρίσκονται σε υψηλό ιεραρχικό επίπεδο. Μια οργανωτική δομή αυτής της μορφής χαρακτηρίζεται ως αυστηρή, γραφειοκρατική, μηχανιστική και επίσημη, ενώ μια δομή με τα ακριβώς αντίθετα χαρακτηριστικά περιγράφεται ως ανεπίσημη ή οργανική. [16]

2.1.4 Στροφή των επιχειρήσεων σε BPM.

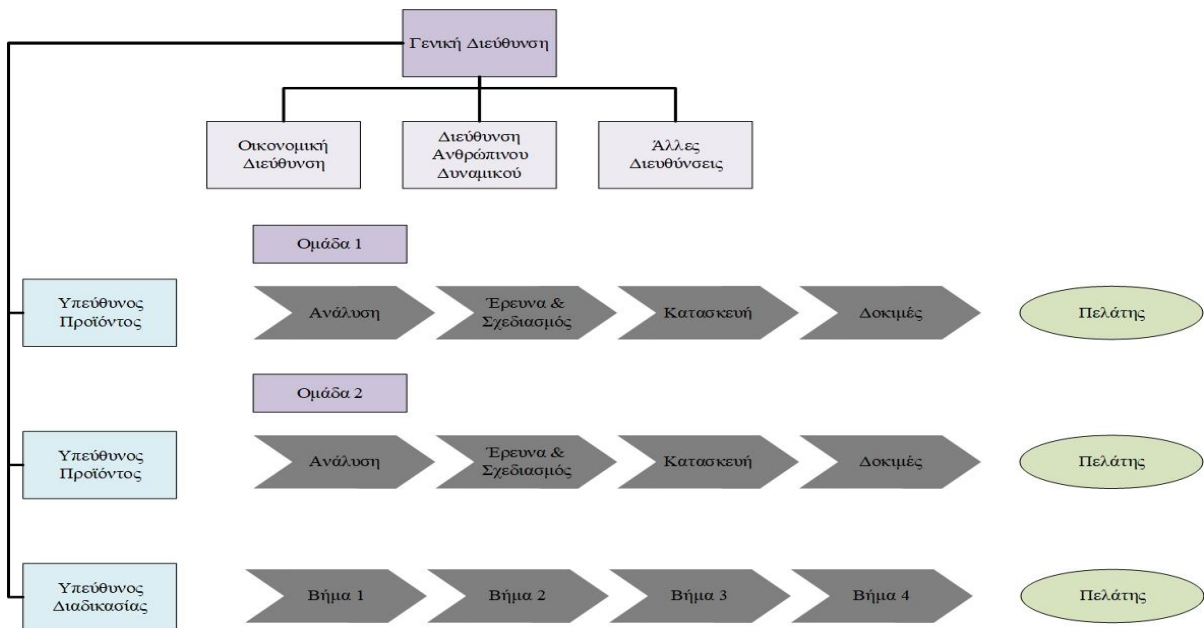
Όπως και είδαμε παραπάνω, το σύνηθες σε έναν οργανισμό είναι μια διαχείριση κάθετα ιεραρχημένη, χωρισμένη σε τμήματα. Αυτό, πολλές φορές δημιουργεί προβλήματα. Ένα από τα προβλήματα αυτά είναι η αναποτελεσματική αντιμετώπιση πολλών διαλειτουργικών ζητημάτων που εμφανίζονται σε έναν οργανισμό. [17]

Υπάρχουν πολλές ιδέες για τη διαχείριση των προβλημάτων που μπορεί να εμφανιστούν με κάποιο οργανόγραμμα παρόμοιο με την εικόνα 1. Στην παρούσα διπλωματική εργασία θα αναφερθούμε στην διαχείριση επιχειρησιακών διαδικασιών.

Επικεντρώνοντας το ενδιαφέρον μας παραπάνω στις διαδικασίες, χρησιμοποιώντας μια οριζόντια διαχείριση διεργασιών ως οργανωτική δομή (εικόνα 3), ο οργανισμός θα θεωρείται ως ένα δίκτυο διεργασιών το οποίο είναι συνδεδεμένο σε ολόκληρο τον οργανισμό.



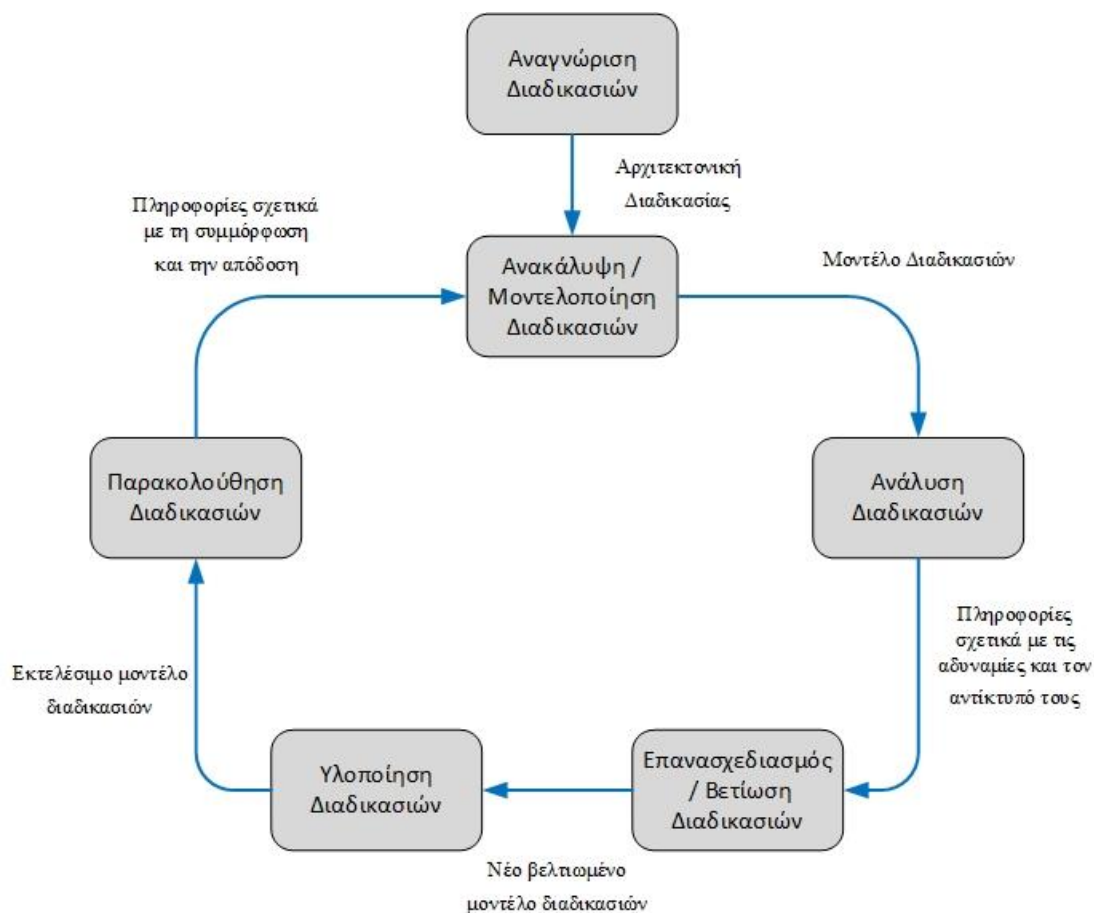
Εικόνα 2. Απεικόνιση των διεργασιών.



Εικόνα 3. Παράδειγμα οριζόντιας οργανωτικής δομής.

2.1.5 Βήματα Διαχείρισης Επιχειρησιακών Διαδικασιών (BPM lifecycle).

Μπορούμε να δούμε την διαχείριση επιχειρησιακών διαδικασιών ως έναν κύκλο, αποτελούμενο από 6 στάδια/φάσεις (εικόνα 4), εξ ου και ο αγγλικός όρος που συναντάμε «BPM lifecycle».



Εικόνα 4. BPM lifecycle.

Ξεκινώντας, έχουμε την αναγνώριση των διαδικασιών (process identification). Σε αυτήν την πρώτη φάση, παρουσιάζεται κάποιο επιχειρησιακό πρόβλημα. Εντοπίζονται όλες οι διαδικασίες που σχετίζονται με αυτό το πρόβλημα και οριοθετούνται. Το αποτέλεσμα της αναγνώρισης των διαδικασιών αυτών είναι μια αρχιτεκτονική, η οποία μας δίνει μια εικόνα σχετικά με τις διαδικασίες που υπάρχουν στον οργανισμό και τις σχέσεις μεταξύ τους. Έπειτα, αυτή η αρχιτεκτονική χρησιμοποιείται ως γνώμονας για να επιλεγθούν οι διαδικασίες που θα περάσουν υπό επεξεργασία στις επόμενες φάσεις του κύκλου. [18]

Δεύτερη σε σειρά φάση είναι η ανακάλυψη διαδικασιών ή και αλλιώς μοντελοποίηση διαδικασιών (process discovery ή και process modeling). Σε αυτό το στάδιο, καταγράφεται η τρέχουσα κατάσταση κάθε διαδικασίας που βρέθηκε στη πρώτη φάση, σε μορφή επιχειρησιακού μοντέλου. [18]

Τρίτη φάση είναι η ανάλυση διαδικασιών (process analysis). Σε αυτήν τη φάση, αναγνωρίζονται όλα τα προβλήματα των διαδικασιών που έχουμε βρει από τα προηγούμενα στάδια και καταγράφονται. Πολλές φορές κατηγοριοποιούνται ανάλογα με την απόδοσή τους. Αυτό που παίρνουμε ως αποτέλεσμα αυτής της φάσης είναι μια συλλογή από προβλήματα. Αυτά τα προβλήματα ταξινομούνται ανάλογα με τον βαθμό επικινδυνότητάς τους και μια εκτίμηση για το πόσο εύκολα μπορούν να λυθούν. [18]

Ακολουθεί η φάση του επανασχεδιασμού διαδικασιών (process redesign). Συνήθως, το στάδιο αυτό ονομάζεται και βελτίωση διαδικασιών (process improvement). Στόχος είναι να βρεθούν αλλαγές στις διαδικασίες που μπορούν να βοηθήσουν να λυθούν τα προβλήματα που εντοπίστηκαν στο παραπάνω στάδιο. Παρουσιάζονται και συγκρίνονται αρκετές αλλαγές αλλά πάντα επικρατεί αυτή με το μεγαλύτερη απόδοση κέρδους. Ως εκ τούτου, οι πιο υποσχόμενες αλλαγές συνδυάζονται με τις διαδικασίες δημιουργώντας έτσι μια νέες επανασχεδιασμένες διαδικασίες και ως αποτέλεσμα της τέταρτης φάσης του κύκλου παίρνουμε ένα νέο μοντέλο διεργασιών. [18]

Πέμπτη φάση είναι η φάση της υλοποίησης διαδικασιών (process implementation). Εδώ, ότι θεωρητικές αλλαγές έχουν γίνει από τα προηγούμενα στάδια ετοιμάζονται και υλοποιούνται. [18]

Τέλος, έκτη και τελευταία, πρόκειται για την φάση της παρακολούθησης των διαδικασιών. Αφού τεθούν σε λειτουργία οι νέες διαδικασίες, συλλέγονται δεδομένα με στόχο την περαιτέρω ανάλυση ώστε να προσδιοριστεί η νέα απόδοση των διαδικασιών και να συγκριθεί με τους στόχους του οργανισμού. Οποιαδήποτε σφάλματα και αποκλίσεις στη συμπεριφορά των διαδικασιών συναντώνται σε αυτή τη φάση διορθώνονται, ώστε όλες οι διαδικασίες να πλησιάζουν τους προβλεπόμενους στόχους. [18]

Οποιαδήποτε νέα προβλήματα προκύπτουν από εδώ και πέρα, επαναλαμβάνεται αυτός ο κύκλος φάσεων.

2.1.6 Πότε πρέπει μια επιχείρηση να ασχοληθεί με το BPM.

Συνήθως η απάντηση διαφέρει ανάλογα με την κατάσταση που βρίσκεται μια επιχείρηση καθώς και με τον βαθμό ωριμότητάς της. Παρακάτω γίνεται μια προσπάθεια να προσδιοριστούν μερικές αιτίες που μπορεί να οδηγήσουν μια επιχείρηση να προβεί σε

BPM. Τα παραδείγματα δεν είναι απόλυτα και δεν καλύπτουν κάθε περίπτωση. Η κάθε επιχείρηση από μόνη της ανάλογα με τον αριθμό των προβλημάτων που θα παρουσιάσει διαλέγει τη στιγμή που θα υλοποιήσει διαχείριση επιχειρησιακών διαδικασιών. [6]

Από τη μεριά του οργανισμού, μερικές περιπτώσεις που συναντάνε και πρέπει να σκεφτούν την υλοποίηση BPM είναι:

- Αδυναμία διαχείρισης υψηλών ρυθμών ανάπτυξης.
- Αναδιοργάνωση οργανισμού.
- Αποτυχία επίτευξης στόχων.
- Ανάγκη για ευελιξία.

Από τη μεριά της διαχείρισης, μερικές περιπτώσεις που συναντάνε και πρέπει να σκεφτούν την υλοποίηση BPM είναι:

- Όταν χρειάζεται να έχουνε παραπάνω έλεγχο των διαδικασιών.
- Ανάγκη χρηματικής απόσβεσης από παλιά συστήματα.
- Χρηματικές περικοπές.
- Ανάγκη δημιουργίας μιας κουλτούρας υψηλής απόδοσης.

Από τη μεριά των εργαζομένων, μερικές περιπτώσεις που συναντάνε και πρέπει η επιχείρηση να σκεφτεί την υλοποίηση BPM είναι:

- Εργαζόμενοι δυσκολεύονται να συμβαδίσουν με τις συνεχείς αλλαγές και την αυξανόμενη πολυπλοκότητα.
- Δυσκολία εκμάθησης εργαζομένων.
- Μη επαρκής έως και καμία ικανοποίηση.
- Επιθυμία να αυξηθεί η ενδυνάμωση των εργαζομένων.

Από τη μεριά των αγοραστών, μερικές περιπτώσεις που συναντάνε και πρέπει η επιχείρηση να σκεφτεί την υλοποίηση BPM είναι:

- Χαμηλή ικανοποίηση με την εξυπηρέτηση
- Μεγάλοι χρόνοι ικανοποίησης παραγγελιών.

Από τη μεριά των διαδικασιών, μερικές περιπτώσεις που συναντάνε και πρέπει η επιχείρηση να σκεφτεί την υλοποίηση BPM είναι:

- Πολλά κενά σε μια διαδικασία, καθόλου σαφής πορεία ενεργειών.
- Έλλειψη σαφών στόχων μιας διαδικασίας.
- Κακή ποιότητα διαδικασιών, πολλές ενέργειες για απλά πράγματα.

[6]

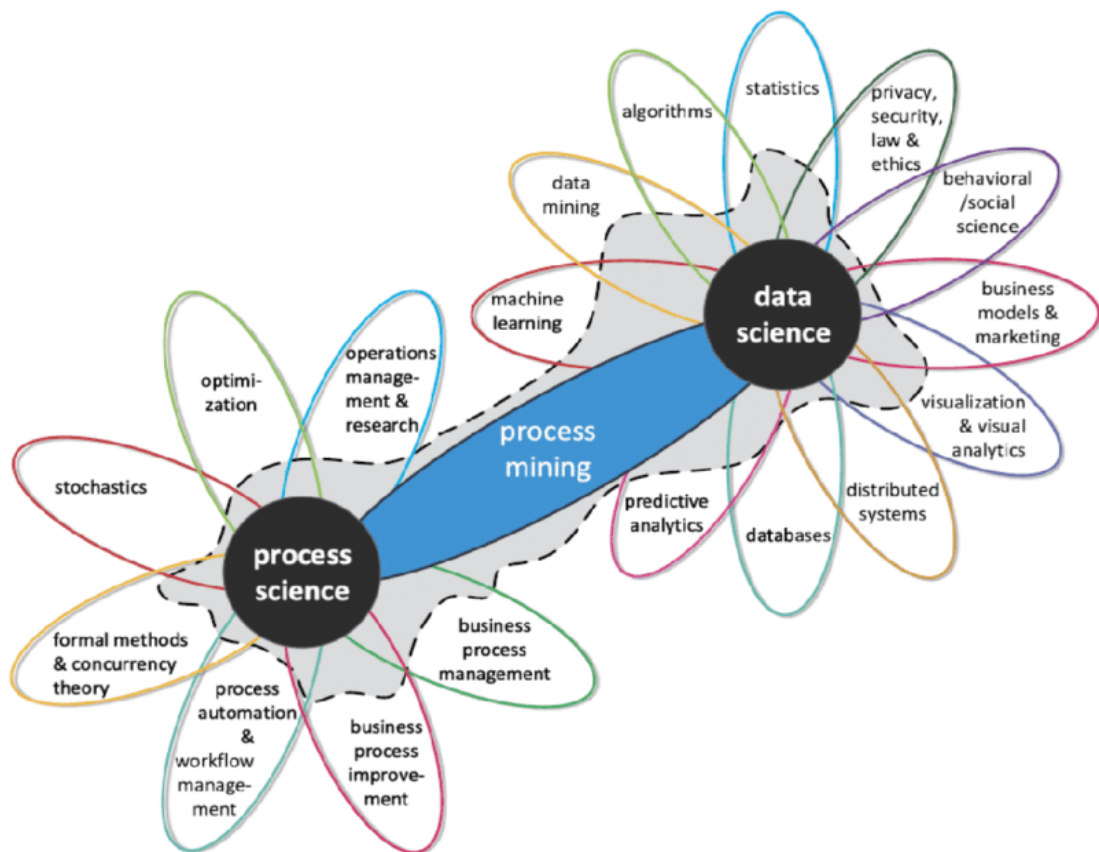
Όταν εμφανιστεί ο «σωστός» αριθμός των παραπάνω προβλημάτων ή και συνδυασμός αυτών, είναι συνετό μια επιχείρηση να προβεί σε διαχείριση επιχειρησιακών διαδικασιών.

2.2 Εξόρυξη Διαδικασιών (Process Mining)

2.2.1 Εισαγωγή στην εξόρυξη διαδικασιών, τύποι, οπτική γωνία και προβλήματα.

2.2.1.1 Εισαγωγή

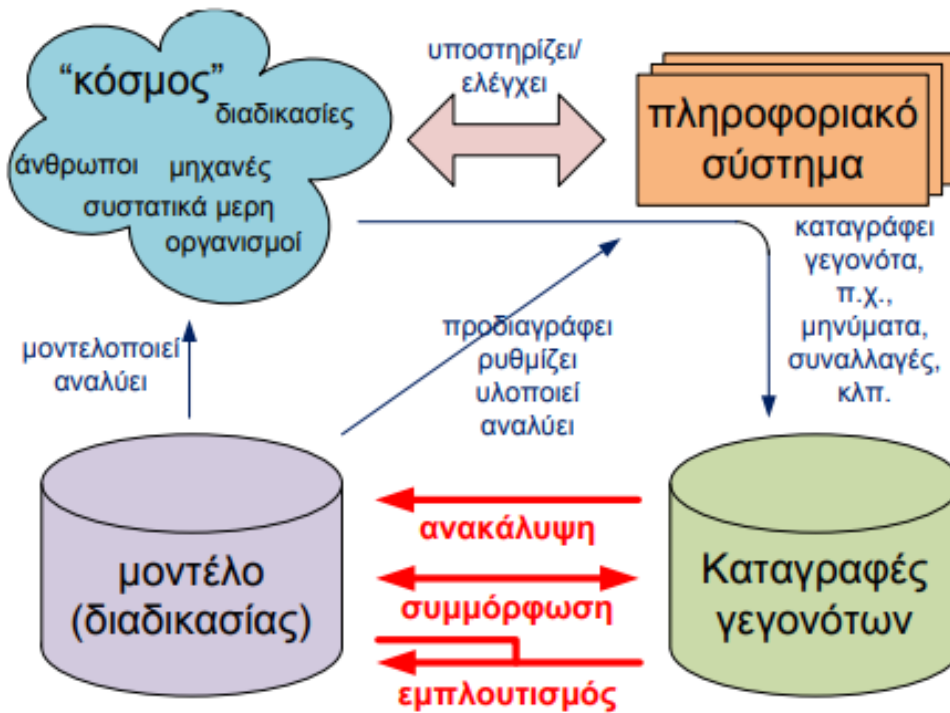
Η εξόρυξη διαδικασιών πρόκειται για έναν νέο σχετικά τομέα έρευνας και βρίσκεται ανάμεσα από τη μηχανική μάθηση και την εξόρυξη δεδομένων από τη μια πλευρά και τη διαχείριση διεργασιών από την άλλη πλευρά (εικόνα 5). [22]



Εικόνα 5. Η εξόρυξη διεργασιών αποτελεί τον σύνδεσμο μεταξύ της επιστήμης δεδομένων και της επιστήμης διεργασιών (W. Van der Aalst, 2016).

Στοχεύει στην ανακάλυψη, παρακολούθηση και βελτίωση των πραγματικών (επιχειρησιακών) διαδικασιών εξάγοντας γνώση από αρχεία καταγραφής συμβάντων, που εύκολα κάθε επιχείρηση μπορεί και παράγει. Περιλαμβάνει αυτοματοποιημένη ανακάλυψη διεργασιών (π.χ. εξαγωγή μοντέλων διαδικασιών από αρχεία καταγραφής

συμβάντων), έλεγχο συμμόρφωσης (π.χ. παρακολούθηση αποκλίσεων συγκρίνοντας το μοντέλο διαδικασιών και το αρχείο καταγραφής συμβάντων/γεγονότων), εξόρυξη κοινωνικού δικτύου, αυτοματοποιημένη κατασκευή μοντέλων προσομοίωσης, πρόβλεψη περιπτώσεων και προτάσεων βάση ιστορικού. [22]



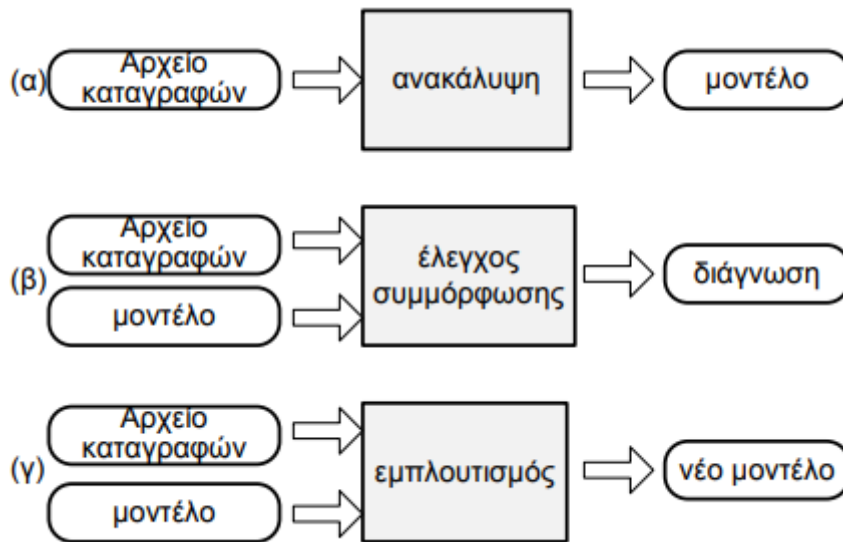
Εικόνα 6. Τα «περιεχόμενα» της εξόρυξης διεργασιών.

Όπως βλέπουμε και στην εικόνα 6, ξεκινάμε με το πληροφοριακό σύστημα, που υποστηρίζει και ελέγχει τις επιχειρηματικές διαδικασίες και αποθηκεύει δεδομένα συμβάντων (π.χ. μηνύματα, συναλλαγές, αρχεία καταγραφής κ.λπ.). Στη συνέχεια, τα δεδομένα αυτά (σε μορφή αρχείων καταγραφής γεγονότων), εξάγονται, φιλτράρονται και φορτώνονται σε κάποιο λογισμικό εξόρυξης διαδικασιών, όπου και διεξάγεται κάποιος τύπος εξόρυξης διαδικασιών (ανακάλυψη, συμμόρφωση, εμπλουτισμός). [22]

2.2.1.2 Τύποι εξόρυξης διεργασιών.

Ένα αρχείο καταγραφής γεγονότων μπορεί περεταίρω να χρησιμοποιηθεί για τη διεξαγωγή τριών τύπων εξόρυξης διεργασιών:

1. ανακάλυψη,
2. συμμόρφωση και
3. εμπλουτισμός.



Εικόνα 7. Οι τρεις τύποι εξόρυξης διαδικασιών σε μορφή εισόδου/εξόδου.

Η παραπάνω εικόνα (εικόνα 7) αναπαριστά τους τύπους της εξόρυξης διαδικασιών σε μορφή εισόδων/εξόδων. Όπως βλέπουμε, αφετηρία για την εξόρυξη διαδικασιών αποτελούν τα αρχεία καταγραφής γεγονότων.

Η τεχνική της ανακάλυψης δέχεται ένα αρχείο καταγραφής γεγονότων και παράγει ένα μοντέλο διαδικασιών. Το παραγμένο μοντέλο είναι συνήθως ένα μοντέλο ροής διαδικασιών (π.χ. ένα διάγραμμα Petri net, BPMN κ.α.). Η τεχνική αυτή είναι η πιο διαδεδομένη τεχνική εξόρυξης δεδομένων. [22]

Η τεχνική του ελέγχου συμμόρφωσης χρειάζεται επιπλέον του αρχείου καταγραφής γεγονότων και ένα πρότυπο μοντέλο ως δεδομένο εισόδου. Συγκρίνεται το αρχείο καταγραφής με το μοντέλο και ελέγχεται εάν η πραγματικότητα, όπως αναφέρεται στο αρχείο καταγραφής, συμμορφώνεται με το μοντέλο και αντίστροφα. Συνήθως χρησιμοποιούμε την τεχνική αυτή για τον έλεγχο της ποιότητας των διαδικασιών, τον εντοπισμό αποκλίσεων στα γεγονότα, τον έλεγχο της ποιότητας ενός μοντέλου και πολλά άλλα. Δύο είναι οι κυριότεροι τύποι ελέγχου συμμόρφωσης που χρησιμοποιούνται σήμερα και θα δούμε και παρακάτω, token-based replay και alignments. Και στις δύο περιπτώσεις ποσοτικοποιείται η έννοια της προσαρμογής (fitness). Στην έξοδο, παραδίδονται

διαγνωστικές αναφορές που περιγράφουν τις διαφορές και τις ομοιότητες μεταξύ του πρότυπου μοντέλου και των καταγραφών. [22]

Η τεχνική του εμπλουτισμού χρειάζεται τα ίδια δεδομένα εισόδου όπως αυτή του ελέγχου συμμόρφωσης (μοντέλο και αρχείο καταγραφής). Στην έξοδο τους παράγουν ένα νέο βελτιωμένο μοντέλο ή ένα μοντέλο στο οποίο έχουν γίνει σημαντικές επεκτάσεις. [9,22]

2.2.1.3 Οπτικές γωνίες εξόρυξης διαδικασιών.

Η εξόρυξη διαδικασιών μπορεί και καλύπτει αρκετές διαφορετικές οπτικές γωνίες. Αρχικά, έχουμε την οπτική της ροής ελέγχου, που επίκεντρο είναι η σειρά των διαδικασιών, η ροή ελέγχου. Στόχος είναι να βρεθεί ένα επιχειρησιακό μοντέλο που να περιγράφει όσο το δυνατόν καλύτερα όλα τα μονοπάτια που μπορεί να πάρει μια διαδικασία. Ακολουθεί η οπτική που ασχολείται με τις διαδικασίες και κύριο μέλημα είναι να αναδειχθεί το πώς εκτελείται μια διαδικασία. Έπειτα, υπάρχει η οπτική γωνία του οργανισμού. Εστιάζει στις πληροφορίες σχετικά με το ποιος εκτέλεσε μια διαδικασία (π.χ. άτομα, επιχειρησιακά τμήματα) και πώς σχετίζονται με αυτήν και συνήθως έτσι δομείται και το κοινωνικό δίκτυο του οργανισμού. Οι δύο τελευταίες οπτικές που καλύπτει η εξόρυξη διαδικασιών είναι η οπτική της περίπτωσης και του χρόνου. Η πρώτη εστιάζει στις ιδιότητες των περιπτώσεων που βρίσκονται στο αρχείο καταγραφής γεγονότων, δηλαδή ασχολείται με την πορεία της διαδικασίας, ποιοι την εκτέλεσαν, ποιο ήταν το κόστος εκτέλεσης κ.α.. Τέλος, η οπτική του χρόνου, ασχολείται με τα χρονοδιαγράμματα μιας διαδικασίας (χρόνος εμφάνισης, συχνότητα διαδικασιών κ.α.). Ουσιαστικά ασχολείται με το χαρακτηριστικό της χρονικής σήμανσης που παρατηρείται σε κάποιο αρχείο καταγραφής γεγονότων και μέσω αυτού είναι εύκολο να ανακαλύψει κανείς σημεία συμφόρησης, να παρακολουθήσει τα επίπεδα εξυπηρέτησης και να προβλέψει τον υπολειπόμενο χρόνο ολοκλήρωσης μιας διαδικασίας που εκτελείται. [22]

2.2.1.4 Συνήθη προβλήματα στην εξόρυξη διαδικασιών

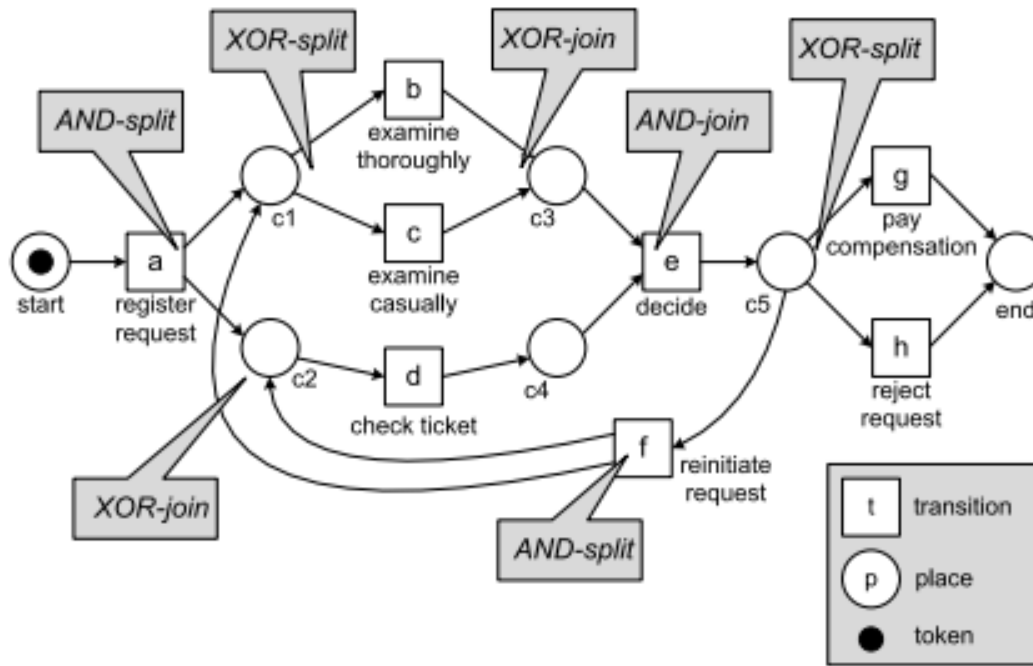
Η εξόρυξη διαδικασιών θεωρείται ένας νέος σχετικά τομέας έρευνας και κατά καιρούς εμφανίζει διάφορα προβλήματα όταν επιχειρείται να εφαρμοστεί σε πραγματικές επιχειρησιακές διαδικασίες. Τα προβλήματα αυτά μπορούν να κατηγοριοποιηθούν σε 10 ομάδες στις οποίες θα γίνει μια συνοπτική αναφορά:

1. Εύρεση, συγχώνευση και φιλτράρισμα δεδομένων αρχείων καταγραφής.
2. Αντιμετώπιση σύνθετων αρχείων καταγραφής γεγονότων με διαφορετικά χαρακτηριστικά.
3. Δυσκολία δημιουργίας αντιπροσωπευτικών μοντέλων αξιολόγησης.
4. Αντιμετώπιση μεταβλητών χαρακτηριστικών των διαδικασιών.
5. Βελτιστοποίηση
6. Ισορροπία μεταξύ των ποιοτικών κριτηρίων αξιολόγησης μοντέλων.
7. Δια-οργανωτική εξόρυξη διαδικασιών.
8. Παροχή υποστήριξης.
9. Συνδυασμός εξόρυξης διαδικασιών με άλλους τύπους ανάλυσης.
10. Βελτίωση κατανόησης και χρηστικότητας από μη ειδικούς.

Μερικά από τα παραπάνω προβλήματα εμφανίστηκαν και κατά την πορεία εξέλιξης της μελέτης περίπτωσης, που μελέτησε η παρούσα διπλωματική και λύθηκαν με την προσέγγιση που ακολουθήθηκε. Προφανώς υπάρχουν διάφορες λύσεις για τις παραπάνω κατηγορίες προβλημάτων και μπορεί να βρει κανείς περισσότερα στο [23]. [22]

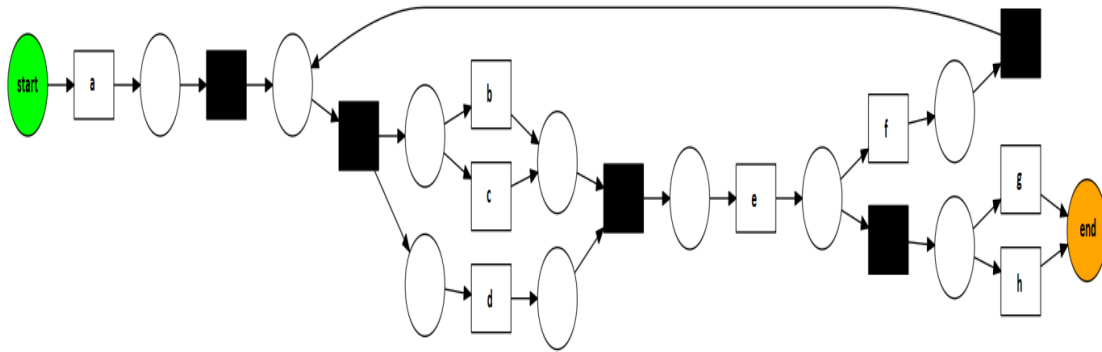
2.2.1.5 Δίκτυα Petri (Petri nets)

Ένα δίκτυο Petri, πρόκειται για μια γραφική απεικόνιση κατάλληλη για μοντελοποίηση συστημάτων και διαδικασιών. Έκανε την εμφάνισή του στις αρχές του 1960 από τον Carl Adam Petri και ήταν ο πρώτος φορμαλισμός που περιέγραφε ικανοποιητικά την ιδιότητα του ταυτοχρονισμού. Πιο αναλυτικά, μιλώντας για ένα Petri net ουσιαστικά μιλάμε για ένα κατευθυνόμενο γράφημα, εμπλουτισμένο με δύο τύπους κόμβων, τα μέρη και τις μεταβάσεις. Οι κόμβοι συνδέονται μεταξύ τους με κατευθυνόμενες ακμές και τα γεωμετρικά σχήματα που χαρακτηρίζουν τους δύο αυτούς κόμβους είναι: κύκλοι για τα μέρη και ορθογώνια παραλληλόγραμμα για τις μεταβάσεις. Δεν επιτρέπονται οι συνδέσεις μεταξύ ιδίου τύπου κόμβων. Ένα απλό παράδειγμα ενός δικτύου Petri φαίνεται παρακάτω στην εικόνα 8. [19]



Εικόνα 8. Παράδειγμα Petri net αποτελούμενο από 8 μέρη και 9 μεταβάσεις. [9]

Οι δραστηριότητες-συμβάντα του αρχείου καταγραφής αντιπροσωπεύονται με τετράγωνα. Αυτές οι μεταβάσεις συνδέονται μεταξύ τους μέσω των κύκλων, δηλαδή κάποιες τοποθεσίες που μοντελοποιούν πιθανές καταστάσεις μιας διαδικασίας. [9, 13] Μερικές φορές σε ένα δίκτυο Petri, μπορεί να παρατηρηθούν και μαύρα χρωματισμένα τετράγωνα, μαύρες μεταβάσεις, που ονομάζονται σιωπηλές ή άορατες μεταβάσεις και δεν αντιπροσωπεύουν καμία δραστηριότητα του αρχείου καταγραφής συμβάντων. Είναι αναγκαίες όμως για τη μοντελοποίηση κάποιας σχέσης μεταξύ των δραστηριοτήτων (σχέσεις παραλληλίας ή αποκλειστικότητας). Αντίστοιχα με AND/OR Gateways σε κάποιο γράφημα για τον προσδιορισμό μιας επιχειρησιακής διαδικασίας (BPMN). Μπορούν να φανούν στην παρακάτω εικόνα.



Εικόνα 9. Εκ νέου παράδειγμα Petri net.

Αν και κατά κανόνα ένα γράφημα Petri net είναι στατικό, τα μέρη ενδέχεται να περιέχουν κάποιες μάρκες (tokens) οι οποίες κινούνται ελεύθερα στο γράφημα, όπου τους επιτρέπεται, σύμφωνα με τον κανόνα πυροδότησης (firing rule). [19]

Σύμφωνα με τον κανόνα πυροδότησης, μια μετάβαση σε ένα Petri net ενεργοποιείται μόνο εάν σε κάθε μέρος εισόδου υπάρχει μια τουλάχιστον μάρκα (token). Κατά την ενεργοποίηση μιας μετάβασης, η μάρκα ή μάρκες καταναλώνονται από τα μέρη εισόδου και προστίθενται στα μέρη-θέσεις εξόδου. Όταν μιλάμε για ενδεικτικό τρέξιμο ενός Petri net, εννοούμε οποιαδήποτε ακολουθία ενεργοποιημένων μεταβάσεων. [13]

Μπορεί κανείς να συναντήσει ένα δίκτυο Petri σε πολλούς τομείς εφαρμογών, για παράδειγμα συστήματα διαχείρισης ροής εργασιών, πρωτόκολλα επικοινωνίας, υπηρεσίες ιστού κ.α. Λόγω των ισχυρών δυνατοτήτων τους, έχουν επηρεάσει και άλλες γλώσσες μοντελοποίησης (UML) αλλά και χρησιμοποιούνται ως βάση πάνω από άλλες γλώσσες για καλύτερη ανάλυση. Τέλος, η εξέλιξη των κλασικών δικτύων Petri ονομάζεται υψηλό επίπεδο δίκτυο Petri και υποστηρίζει την μοντελοποίηση περισσότερο πολύπλοκων συστημάτων αλλά και επιτρέπει καλύτερη απεικόνιση πληροφοριών στο γράφημα. [19]

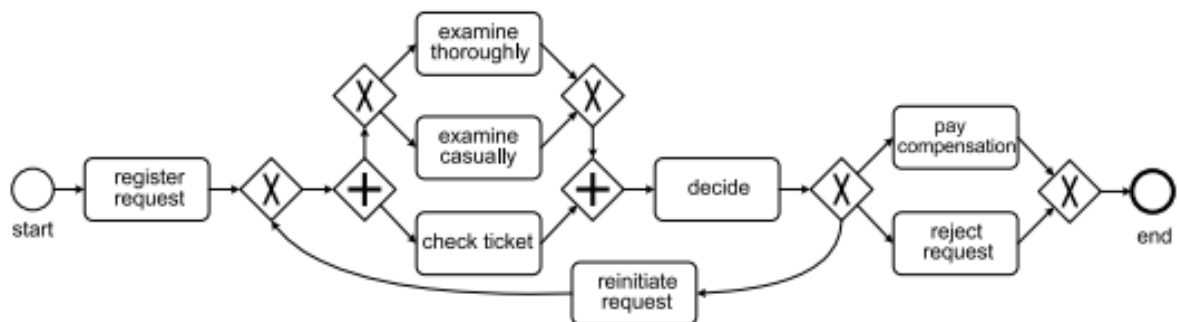
2.2.1.6 Μοντελοποίηση επιχειρησιακών διαδικασιών (BPMN)

Το Business Process Modeling Notation (BPMN) αποτελεί ένα γραφικό συμβολισμό για τη σχεδίαση επιχειρησιακών διαδικασιών και είναι η πλέον δημοφιλέστερη γλώσσα μοντελοποίησης επιχειρησιακών διαδικασιών. Η γλώσσα είναι παρόμοια με άλλες γνωστές γλώσσες που χρησιμοποιούν συμβολισμούς (UML) και εκτεταμένες αλυσίδες διεργασιών. Αν και αναπτύχθηκε από το Business Process Management Initiative (BPMI),

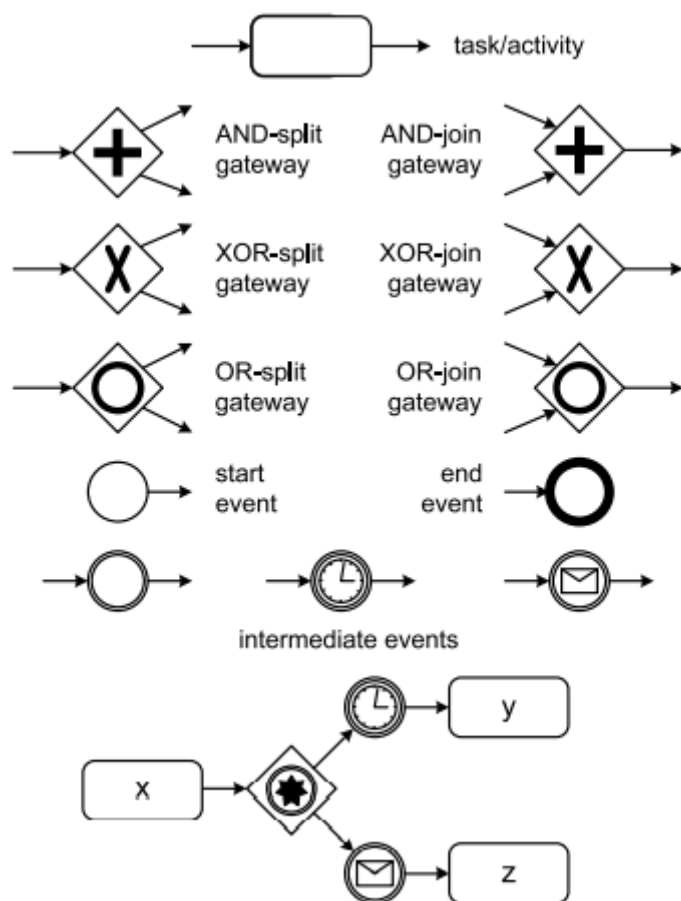
πλέον συντηρείται από το Object Management Group (OMG) και στοχεύει στο να τυποποιήσει μια επιχειρησιακή μοντελοποίηση διαδικασιών έναντι όλων των άλλων που υπάρχουν. [20]

Όταν κάποιος μιλάει για ένα μοντέλο BPMN, αναφέρεται σε κάποιο διάγραμμα επιχειρησιακής διαδικασίας (BPD). Ένα τέτοιο διάγραμμα είναι ουσιαστικά ένα διάγραμμα ροής που αποτελείται από πολλά διαφορετικά στοιχεία. Υπάρχουν 4 κατηγορίες στοιχείων, τα flow objects, τα connecting objects, τα swimlanes και τα artifacts. Η πρώτη κατηγορία περιέχει τα βασικά γραφικά στοιχεία που προσδιορίζουν τη συμπεριφορά μιας διεργασίας. Χωρίζονται σε events, activities και gateways. Τα event είναι ότι και τα μέρη στα δίκτυα Petri που είδαμε παραπάνω, τα activities συχνά αναφέρονται ως tasks και πρόκειται για τις δραστηριότητες-συμβάντα και τα gateways ή πύλες χρησιμοποιούνται για να δημιουργήσουν ενώσεις και διαχωρισμούς. Τα στοιχεία αυτά συνδέονται μεταξύ τους μέσω των connecting objects και δημιουργούνται σχέσεις και ροή ελέγχου. Η τρίτη κατηγορία (swimlanes) απλώς βοηθάει τη σύνθεση δομής των διαδικασιών ενώ η τελευταία κατηγορία (artifacts) χρησιμοποιείται κυρίως για περαιτέρω σχολιασμό των μοντέλων. [20]

Στην εικόνα 10 παρατηρούμε ένα επιχειρησιακό μοντέλο σχεδιασμένο με BPMN, ενώ η εικόνα 11 δείχνει τα βασικά σύμβολα που χρησιμοποιεί η γλώσσα BPMN. Αν και υπάρχουν πάνω από 50 διαφορετικά σύμβολα, έχει αποδειχθεί ότι ένα μέσο υποσύνολο ενός BPMN μοντέλου χρησιμοποιεί λιγότερα από 10 διαφορετικά σύμβολα. Ένας απλός συνδυασμός των συμβόλων της εικόνας 11 αρκεί για την κατασκευή ενός ισχυρού διαγράμματος. [9]



Εικόνα 10. Παράδειγμα διαγράμματος BPMN. [9]



Εικόνα 11. Σύνολο συμβόλων που χρησιμοποιεί το BPMN. [9]

2.2.2 Αρχεία καταγραφής γεγονότων, μοντέλα διεργασιών και οι σχέσεις μεταξύ τους (Play-Out, Play-In , Replay).

2.2.2.1 Αρχεία καταγραφής γεγονότων – event logs

Αφετηρία για την εξόρυξη διαδικασιών αποτελούν τα αρχεία καταγραφής γεγονότων. Ένα αρχείο καταγραφής γεγονότων συγκροτείται από ένα πλήθος περιπτώσεων, όπου κάθε περίπτωση είναι μια μοναδική περίπτωση εκτέλεσης μιας διαδικασίας. Κάθε περίπτωση αποτελείται από μια ακολουθία γεγονότων, τα οποία γεγονότα όντας το βασικό συστατικό των αρχείων καταγραφής, χαρακτηρίζονται από ένα σύνολο χαρακτηριστικών. Μερικά από τα στοιχεία που τα χαρακτηρίζουν είναι τα εξής: i) χρονική σήμανση, ii) όνομα δραστηριότητας, iii) χρονοδιάγραμμα κ.α.. [9]

Ορισμός (αρχείο καταγραφής γεγονότων): Ένα αρχείο καταγραφής γεγονότων πρόκειται για ένα σύνολο (Γ, ΤΓ, ΕΓ, Π, χρόνος, ετικέτα, τύπος, περίπτωση), όπου Γ ορίζουμε ένα σύνολο γεγονότων, ΤΓ= (εκκίνηση, ολοκλήρωση) ορίζουμε ένα σύνολο τύπων γεγονότων, ΕΓ ορίζουμε ένα σύνολο ετικετών των γεγονότων, Π ορίζουμε το σύνολο των περιπτώσεων, χρόνος: $\Gamma \rightarrow \mathbb{R}^+$ ορίζουμε μια συνάρτηση που εκχωρεί μια χρονική σήμανση σε κάποιο γεγονός, ετικέτα: $\Gamma \rightarrow \text{ΕΓ}$ ορίζουμε μια συνάρτηση που εκχωρεί μια ετικέτα δραστηριότητας σε κάποιο γεγονός, τύπος: $\Gamma \rightarrow \text{ΤΓ}$ ορίζουμε μια συνάρτηση που εκχωρεί ένα τύπο γεγονότος σε ένα γεγονός και περίπτωση: $\Gamma \rightarrow \text{Π}$ ορίζουμε μια συνάρτηση που συνδέει ένα γεγονός με μια περίπτωση. [9]

Case ID	Time stamp	Activity	Resource	Cost
1	30-12-2010:11.02	Registration	Peter	50
1	30-12-2010:12.08	IV Antibiotics	Maria	60
1	30-12-2010:13.16	Sepsis triage	Anna	100
2	30-12-2010:16.03	Registration	Peter	50
2	30-12-2010:17.52	Sepsis triage	Anna	150
3	30-12-2010:17.57	Registration	George	55
3	30-12-2010:18.19	Sepsis triage	Anna	145

Εικόνα 12. Παράδειγμα μέρους αρχείου καταγραφής γεγονότων.

Οι πιο διαδεδομένες μορφές αρχείων καταγραφής γεγονότων είναι οι εξής: XES (eXtensible Event Stream), MXML (Mining eXtensible Markup Language) και αρχεία CSV (comma-separated values ή υπολογιστικό φύλλο). Από το 2010 μέχρι και σήμερα, το XES έχει υιοθετηθεί ως η πρότυπη μορφή για τη καταγραφή γεγονότων από την ομάδα IEEE (Institute of Electrical and Electronics Engineers) και υποστηρίζεται από την πλειοψηφία των εργαλείων εξόρυξης δεδομένων. [9]

```

<?xml version="1.0" encoding="UTF-8" ?>
  <extension name="Concept" prefix="concept" uri="http://.../concept.xesext"/>
  <extension name="Time" prefix="time" uri="http://.../time.xesext"/>
  <extension name="Organizational" prefix="org" uri="http://.../org.xesext"/>
  <global scope="trace">
    <string key="concept:name" value="name"/>
  </global>
  <global scope="event">
    <date key="time:timestamp" value="2010-12-17T20:01:02.229+02:00"/>
    <string key="concept:name" value="name"/>
    <string key="org:resource" value="resource"/>
  </global>
  <classifier name="Activity" keys="concept:name"/>
  <classifier name="Resource" keys="org:resource"/>
  <classifier name="Both" keys="concept:name org:resource"/>
  <trace>
    <string key="concept:name" value="1"/>
    <event>
      <string key="concept:name" value="register request"/>
      <string key="org:resource" value="Pete"/>
      <date key="time:timestamp" value="2010-12-30T11:02:00.000+01:00"/>
      <string key="Event_ID" value="35654423"/>
      <string key="Costs" value="50"/>
    </event>
    <event>
      <string key="concept:name" value="examine thoroughly"/>
      <string key="org:resource" value="Sue"/>
      <date key="time:timestamp" value="2010-12-31T10:06:00.000+01:00"/>
      <string key="Event_ID" value="35654424"/>
      <string key="Costs" value="400"/>
    </event>
    <event>
      <string key="concept:name" value="check ticket"/>
      <string key="org:resource" value="Mike"/>
      <date key="time:timestamp" value="2011-01-05T15:12:00.000+01:00"/>
      <string key="Event_ID" value="35654425"/>
      <string key="Costs" value="100"/>
    </event>
    <event>
      <string key="concept:name" value="decide"/>
      <string key="org:resource" value="Sara"/>
      <date key="time:timestamp" value="2011-01-06T11:18:00.000+01:00"/>
      <string key="Event_ID" value="35654426"/>
      <string key="Costs" value="200"/>
    </event>
    <event>
      <string key="concept:name" value="reject request"/>
      <string key="org:resource" value="Pete"/>
      <date key="time:timestamp" value="2011-01-07T14:24:00.000+01:00"/>
      <string key="Event_ID" value="35654427"/>
      <string key="Costs" value="200"/>
    </event>
  </trace>
  <trace>
    <string key="concept:name" value="2"/>
    <event>
      <string key="concept:name" value="register request"/>
      <string key="org:resource" value="Mike"/>
      <date key="time:timestamp" value="2010-12-30T11:32:00.000+01:00"/>
      <string key="Event_ID" value="35654483"/>
      <string key="Costs" value="50"/>
    </event>
    ...
  </trace>
  ...
</log>

```

Εικόνα 13. Παράδειγμα μέρους XES αρχείου.

Για να θεωρηθεί ένα αρχείο καταγραφής γεγονότων έγκυρο, θα πρέπει να εμπεριέχει τουλάχιστον ένα αναγνωριστικό της περίπτωσης, το όνομα των δραστηριοτήτων και τη χρονική σήμανση. Επιπλέον χαρακτηριστικά που μπορεί να εμπεριέχονται, συνήθως περιγράφουν είτε το γεγονός είτε την περίπτωση και δεν τα συναντάμε πάντα σε κάθε αρχείο καταγραφής γεγονότων. Παραδείγματος χάριν, στην εικόνα 8 αν κοιτάξουμε τις δύο τελευταίες στήλες, μπορούμε να διακρίνουμε από δύο επιπλέον χαρακτηριστικό προς το γεγονός, το ανθρώπινο δυναμικό και το κόστος.

Δεν υπάρχει συγκεκριμένο μέγεθος για ένα αρχείο καταγραφής συμβάντων. Αυτό σημαίνει ότι ανάλογα με τις απαιτήσεις του ο καθένας διαλέγει ένα αντιπροσωπευτικό δείγμα που πιστεύει ότι θα είναι επαρκές. Βέβαια, αν κύριο μέλημα κάποιου είναι η κατανόηση μιας απλής, μικρής, τυπικής διαδικασίας, έχοντας ένα μεγάλο πλήθος δεδομένων δεν θα του προσφέρει τίποτα παραπάνω από ένα πιο μικρό δείγμα. Σε αντίθεση με κάποιον άλλο που χρειάζεται, για παράδειγμα, να αναζητήσει παρατυπίες ή λάθη από μεριάς συμμόρφωσης, θα ήταν φρόνιμο να ελέγξει όσο πιο πολλά δεδομένα γίνεται, την ανάλογη ελεγχόμενη περίοδο. Αν και συνήθως επικρατεί ο άγραφος κανόνας «όσο περισσότερα τα δεδομένα τόσο το καλύτερο». Ένα θετικό που έχει η εξόρυξη διεργασιών σε σχέση με τεχνικές εξόρυξης δεδομένων ή και άλλων τεχνικών στατιστικής είναι ότι δεν υπάρχει ελάχιστος αριθμός δεδομένων σε ένα αρχείο καταγραφής συμβάντων. Μπορεί κανείς να χρησιμοποιήσει αρχείο καταγραφής συμβάντων με τρεις περιπτώσεις και να δει αναλυτικά τις διαδρομές που ακολούθησαν αυτές οι περιπτώσεις. [9]

2.2.2.2 Μοντέλα διεργασιών – process models

Ένα μοντέλο διεργασιών, αναπαριστά γραφικά μια ροή εργασίας ή διεργασιών. Ουσιαστικά μας επιτρέπει την οπτικοποίηση των διεργασιών (μιας επιχείρησης) με στόχο την παραπάνω κατανόηση και διαχείριση. Συνήθως, χρησιμοποιούνται για την παρακολούθηση των σχέσεων των διεργασιών κατά τη διάρκεια μιας διαδικασίας. Οι επιχειρήσεις με αυτόν τον τρόπο μπορούν να παρατηρήσουν πως εκτελούνται οι διεργασίες και ποια μέρη χρίζουν βελτίωση αλλά και μέσω κανόνων να δημιουργήσουν μια ιδανική για αυτούς, όσον αφορά την απόδοση, ροή διεργασιών. Η πιο γνωστή γλώσσα αναπαράστασης τέτοιων μοντέλων είναι η BPMN (Business Process Model and Notation). Μια ακόμη μορφή μοντέλου διεργασιών είναι τα δίκτυα Petri (Petri nets).

2.2.2.3 Σχέσεις μεταξύ αρχείου καταγραφής συμβάντων και μοντέλου διεργασιών.

Ένα από τα σημαντικότερα πράγματα στην εξόρυξη διεργασιών είναι να καταφέρουμε να δημιουργήσουμε μια δυνατή και υγιής σχέση μεταξύ μοντέλου διεργασιών και της πραγματικότητας, δηλαδή αυτού που βλέπουμε μέσα από το αρχείο καταγραφής γεγονότων. Έχουν επικρατήσει τρεις έννοιες που περιγράφουν αυτήν τη σχέση: Play-Out, Play-In και Replay. (εικόνα 14) [9]

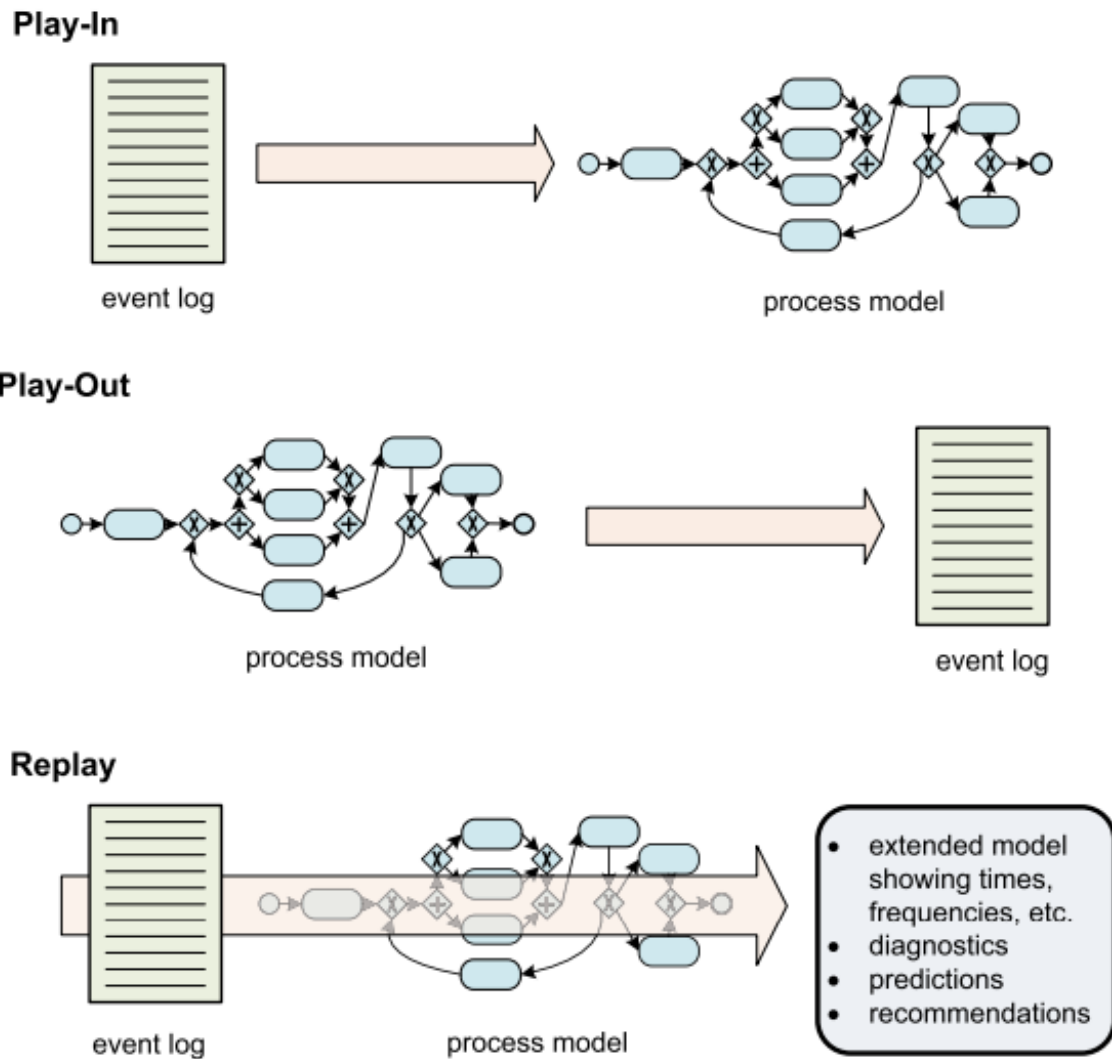
Ξεκινώντας με το Play-Out, έχουμε την κλασική λειτουργία ενός μοντέλου διεργασιών. Δεδομένου ενός γραφήματος, εξάγουμε συμπεριφορά. Στόχος του Play-Out είναι να επαναλαμβάνεται η ροή των διεργασιών που βρίσκονται σε ένα γράφημα ώστε στο τέλος να μαζέψουμε τη σειρά των εκτελέσιμων, αλλά και μη, διαδικασιών. Χρησιμοποιείται αρκετά για ανάλυση αλλά και για αναπαράσταση επιχειρησιακών διαδικασιών. Αρκετά εργαλεία προσομοίωσης χρησιμοποιούν το Play-Out για τη διεξαγωγή πειραμάτων πάνω στις διεργασίες αλλά και να συλλέγουν στατιστικά στοιχεία πάνω σε αυτές. [9]

Το Play-In, είναι το ακριβώς αντίθετο του Play-Out. Αντί να παίρνουμε ένα μοντέλο διεργασιών και να δημιουργούμε το αρχείο καταγραφής συμβάντων, παίρνει ως είσοδο συμπεριφορές, τις εξετάζει και δημιουργεί ένα μοντέλο διεργασιών. Ουσιαστικά πρόκειται για τη δημιουργία ενός μοντέλου διεργασιών μέσω ενός αρχείου καταγραφής συμβάντων. Οι αλγόριθμοι ανακάλυψης μοντέλων διεργασιών που θα συναντήσουμε στα κεφάλαια 3 και 4 πρόκειται για τεχνικές Play-In.

Τέλος, το Replay, αναπαράγει το αρχείο καταγραφής συμβάντων πάνω από ένα μοντέλο διεργασιών. Σκοπός αυτής της διαδικασίας είναι είτε:

- να βρεθούν διαφορές, αποκλίσεις μεταξύ αρχείου καταγραφής συμβάντων και μοντέλου διεργασιών,
- να βρεθούν σημεία συμφόρησης,
- να εμπλουτιστεί το μοντέλο με χρονικές πληροφορίες (βλέποντας ποιες διεργασίες επισκεπτόντουσαν πιο συχνά),
- να δημιουργηθεί ένα νέο μοντέλο πρόβλεψης (με την επανάληψη του αρχείου καταγραφής συμβάντων πολλές φορές), κ.α.

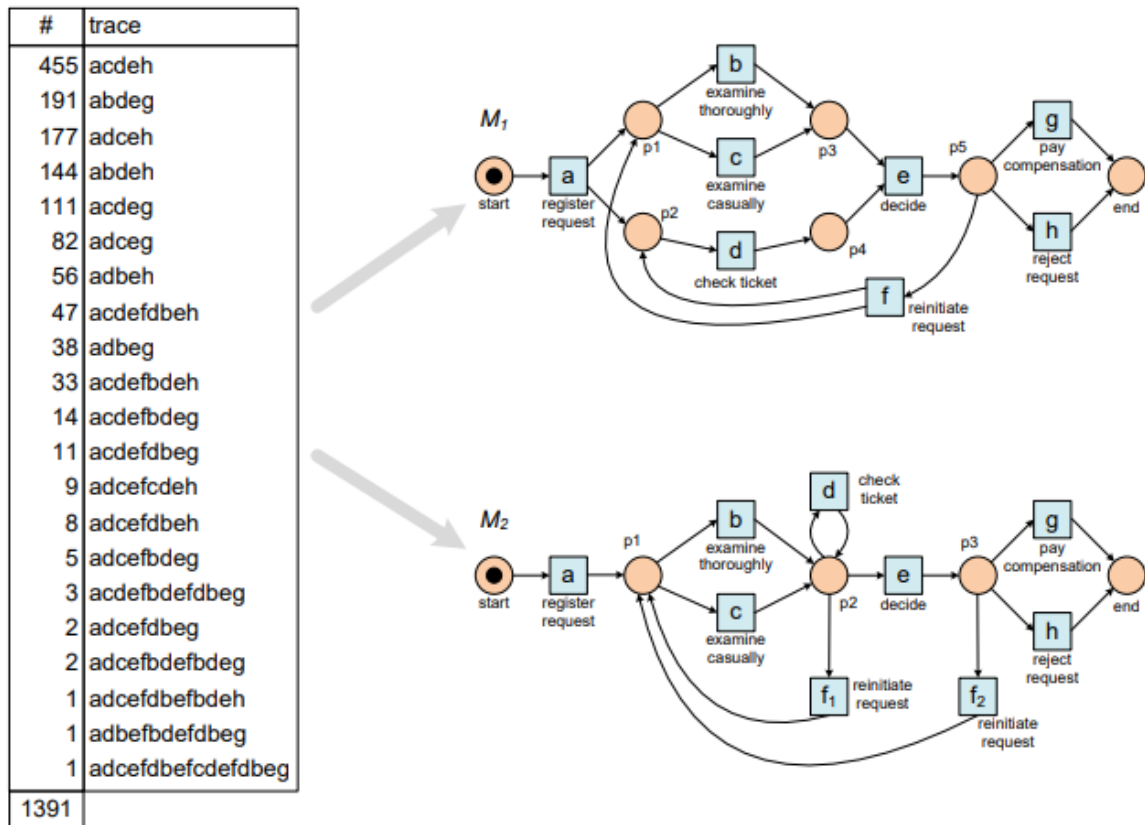
[9].



Εικόνα 14. Play-In, Play-Out και Replay. [9]

2.2.3 Ανακάλυψη.

Με την τεχνική της ανακάλυψης παράγεται αυτόματα ένα μοντέλο διαδικασιών από κάποιο αρχείο καταγραφής γεγονότων, χωρίς καμία επιπλέον πληροφορία. Πρόκειται για την πιο σημαντική λειτουργία της εξόρυξης διαδικασιών. [9,24]



Εικόνα 15. Στα αριστερά απεικονίζεται ένα αρχείο καταγραφής γεγονότων και στα δεξιά απεικονίζονται δύο ενδεικτικά μοντέλα διαδικασιών που περιγράφουν τη συμπεριφορά που παρατηρήθηκε στο αρχείο. [24]

2.2.3.1 Περιπτώσεις εφαρμογής της τεχνικής ανακάλυψης διαδικασιών.

Επικρατεί σήμερα στους οργανισμούς να λειτουργούν με μια σειρά διαδικασιών όταν θέλουν να διαχειριστούν κάποια περίπτωση. Αν και τις περισσότερες φορές αυτές οι διαδικασίες είναι απαραίτητες και συνίστανται από το πληροφοριακό σύστημα του οργανισμού, συνήθως είναι άτυπες και δεν καταχωρούνται πουθενά. Και αν τύχει και καταχωρηθούν, μπορεί να διαφέρουν από την πραγματικότητα. Επομένως, αξίζει να

ανακαλύψει κανείς τις πραγματικές διαδικασίες μέσω των καταγεγραμμένων γεγονότων. Τα παραγόμενα επιχειρησιακά μοντέλα διαδικασιών, μπορούν να χρησιμοποιηθούν για:

1. Τη συζήτηση προβλημάτων μεταξύ των ενδιαφερόμενων μέσα στον οργανισμό (χρειάζεται να γνωρίζουν πραγματικά ποιες διαδικασίες γίνονται, ώστε να έχουν μια κοινή άποψη για το τι συμβαίνει),
2. Τη δημιουργία ιδεών βελτίωσης των διαδικασιών (βλέποντας τις πραγματικές διαδικασίες και τα προβλήματα που παρουσιάζουν, γίνονται προσπάθειες λύσης αυτών),
3. Τη βελτίωση κάποιου επιχειρησιακού μοντέλου (μέσω ανάλυσης),
4. Τη διαμόρφωση (και εκ νέου) ενός συστήματος BPM (το παραγόμενο μοντέλο χρησιμοποιείται ως πρότυπο).

[24]

2.2.3.2 Αλγόριθμοι ανακάλυψης μοντέλων διαδικασιών.

Οι πιο γνωστοί αλγόριθμοι ανακάλυψης μοντέλου διαδικασιών είναι οι:

1. Alpha,
2. Heuristic και
3. Inductive.

Ο αλγόριθμος alpha, είναι ο πιο γνωστός αλγόριθμος ανακάλυψης ενός μοντέλου διαδικασιών. Ψάχνει το αρχείο καταγραφής γεγονότων για διάφορα μοτίβα, σχέσεις μεταξύ των δραστηριοτήτων και δημιουργεί «σχεσιακές εξαρτήσεις» μεταξύ αυτών. Έτσι, καταφέρνει και σχεδιάζει ένα δίκτυο Petri με τους γνωστούς κόμβους και ακμές, που είδαμε παραπάνω στην ενότητα 2.2.3. Συνεχίζοντας, έχουμε τον heuristic αλγόριθμο. Πρόκειται για μια βελτιωμένη μορφή του αλγορίθμου alpha και διαχειρίζεται πλέον το θόρυβο. Βασίζεται αρκετά στις συχνότητες εμφάνισης των δραστηριοτήτων και καταφέρνει να αναπαριστά τους κύκλους στο αρχείο καταγραφής γεγονότων. Τέλος, ο αλγόριθμος inductive είναι μια βελτιωμένη μορφή των alpha και heuristics. Εγγυάται ένα μοντέλο με αρκετά υψηλή τιμή προσαρμογής (fitness). Ουσιαστικά χωρίζει σε κομμάτια μια περίπτωση και προσπαθεί να καταλάβει τις σχέσεις ένωσης, παραλληλίας, αποκλειστικότητας κλπ. μεταξύ των κομματιών αυτών. Μέσω του αλγορίθμου αυτού μπορούμε εύκολα να παράξουμε και το γνωστό δένδρο διαδικασιών. [9,24,25]

Συνοπτικά, μια σύγκριση μεταξύ των αλγορίθμων αυτών θα μπορούσε να δείχνει ως εξής:

Αλγόριθμος Alpha:

- Δεν διαχειρίζεται σωστά τυχόν κύκλους που μπορεί να υπάρχουν στο αρχείο καταγραφής γεγονότων,
- αόρατες και διπλότυπες δραστηριότητες δεν μπορούν να ανακαλυφθούν,
- το μοντέλο μπορεί να μην είναι 100% ορθό, έχει «ήχο»,
- το μοντέλο είναι «αδύναμο» όσον αφορά το θόρυβο του αρχείου καταγραφής

Αλγόριθμος Heuristic

- Λαμβάνει υπόψη τη συχνότητα των δραστηριοτήτων,
- ανακαλύπτει κύκλους,
- το μοντέλο μπορεί να μην είναι 100% ορθό, έχει «ήχο».

Αλγόριθμος Inductive

- Μπορεί να διαχειριστεί αόρατες δραστηριότητες
- είναι ο πιο χρησιμοποιούμενος αλγόριθμος ανακάλυψης μοντέλων διαδικασιών
- το μοντέλο είναι 100% ορθό.

[9,24,25]

2.2.4 Κριτήρια αξιολόγησης ενός μοντέλου διαδικασιών.

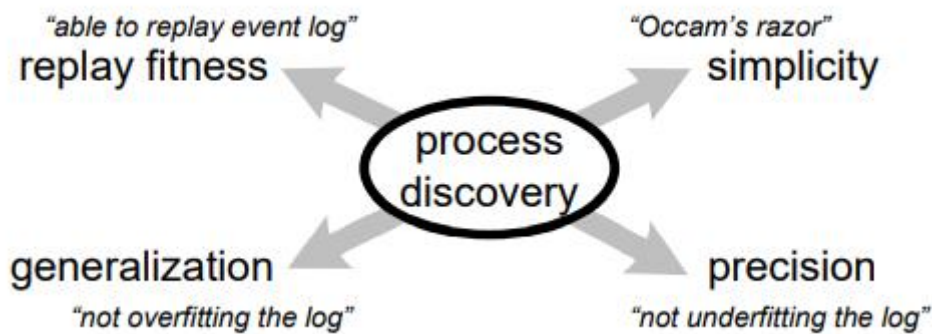
Όπως είδαμε παραπάνω, αρκετοί είναι οι αλγόριθμοι που, μέσω ενός αρχείου καταγραφής συμβάντων, μπορούν και παράγουν ένα μοντέλο διαδικασιών με στόχο την ακριβή περιγραφή των διαδικασιών που λαμβάνουν τόπο σε μια επιχείρηση. Μιας και τα μοντέλα παρατηρούνται σε διάφορες μορφές (Petri net, BPMN κ.α.) έχουν γίνει προσπάθειες για τη σύγκριση και την αξιοπιστία τους βασιζόμενοι στις ιδιότητές τους. [21]

Γενικά, όταν μιλάμε για την ποιότητα ενός παραγόμενου μοντέλου διαδικασιών, τέσσερις διαστάσεις ποιότητας κρίνουν το αποτέλεσμα. Φυσικά, πρόκειται για:

1. Προσαρμογή (fitness / replay fitness)

2. απλότητα (simplicity),
3. ακρίβεια (precision) και
4. γενίκευση (generalization).

Και αυτό γίνεται διότι ένα μοντέλο πολλές φορές μπορεί να μας προσφέρει παραπάνω πληροφορίες από ότι ένα αρχείο καταγραφής συμβάντων. Δηλαδή, λόγω κάποιου βρόγχου επανάληψης στο αρχείο καταγραφής, το μοντέλο να δείχνει συμβάντα που μπορεί να μην υπάρχουν στο αρχείο, ή να υπάρχουν σε πολύ μικρό βαθμό είτε να θεωρούνται ως θόρυβος. Επομένως, είναι σκόπιμο να μετράμε την ποιότητα του μοντέλου από διαφορετικές αντιλήψεις. (εικόνα 16)



Εικόνα 16. Διαστάσεις ποιότητας μοντέλου διαδικασιών. [21]

Ξεκινώντας με την προσαρμογή (replay fitness), η διάσταση αυτή περιγράφει το ποσοστό της συμπεριφοράς που παρατηρείται στο αρχείο καταγραφής συμβάντων και μπορεί να αναπαραχθεί από το μοντέλο. Όσο ψηλότερο είναι το ποσοστό αυτό, τόσο περισσότερα συμβάντα του αρχείου καταγραφής μπορούν να αναπαραχθούν από το μοντέλο. Όταν ο αριθμός της διάστασης αυτής είναι 1, τότε λέμε ότι το μοντέλο παρουσιάζει τέλεια προσαρμογή και όλα τα συμβάντα που παρατηρούνται στο αρχείο καταγραφής μπορούν να αναπαραχθούν από το μοντέλο από την αρχή ως το τέλος. [21]

Η διάσταση της απλότητας (simplicity), περιγράφει το ποσοστό απλότητας του μοντέλου. Δηλαδή πόσο απλό είναι το μοντέλο στο να κατανοηθεί από τον άνθρωπο. Αν και η διάσταση αυτή δεν συνδέεται άμεσα με το αρχείο καταγραφής συμβάντων, είναι απαραίτητη καθώς εξετάζει το μοντέλο καθαυτού. Δεδομένου ότι υπάρχουν αρκετοί

τρόποι για να μετρηθεί η συμπεριφορά από διαφορετικά μοντέλα, η επιλογή του απλούστερου μοντέλου φαντάζει καλύτερη. Η αρχή αυτή είναι γνωστή και ως «λεπίδα του Όκαμ» (Occam's razor): “δεν πρέπει να αυξηθεί, πέραν από αυτό που είναι απαραίτητο, ο αριθμός των οντοτήτων που απαιτούνται για να εξηγήσουν οτιδήποτε.” Τέλος, έρευνες έχουν δείξει ότι ένας σημαντικός παράγοντας που επηρεάζει την απλότητα ενός μοντέλου είναι το μέγεθος και ότι ένα πολύπλοκο μοντέλο συνήθως αποτελείται από ένα μεγάλο αριθμό διαδικασιών. [21]

Η ακρίβεια (precision), περιγράφει το μέγεθος της συμπεριφοράς που επιτρέπει ένα μοντέλο διαδικασιών και δεν παρατηρείται στο αρχείο καταγραφής συμβάντων. Ένα μοντέλο καλείται ακριβές όταν δεν επιτρέπει περισσιες συμπεριφορές, παρά μόνο του αρχείου καταγραφής. [21]

Η προσαρμογή, η απλότητα και η ακρίβεια από μόνες τους δεν επαρκούν για να κρίνουν την ποιότητα ενός μοντέλου διαδικασίας που ανακαλύπτεται. Γενικά, μπορεί να κατασκευαστεί ένα πολύ απλό μοντέλο ικανό να επιτρέπει την αναπαραγωγή όλων των συμβάντων του αρχείου καταγραφής χωρίς όμως να είναι και απόλυτα ακριβές, για παράδειγμα δεν έχουν καταγραφεί όλα τα πιθανά σενάρια βρόγχων επανάληψης στο αρχείο καταγραφής ή επιτρέπει επιπλέον συμπεριφορές διαφορετικές του αρχείου καταγραφής. Οι δύο αυτές περιπτώσεις ονομάζονται υπερπροσαρμογή και υποπροσαρμογή αντίστοιχα. Έτσι, έχουμε και τη διάσταση της γενίκευσης (generalization) που περιγράφει αυτό το φαινόμενο. Ένα μοντέλο, συνίσταται να γενικεύει και να μην περιορίζει τις συμπεριφορές που παρατηρούνται στα συμβάντα του αρχείου καταγραφής. [21]

Γενικά, είναι μια πρόκληση να βρεθεί μια ισορροπία μεταξύ των τεσσάρων αυτών διαστάσεων ποιότητας. Οι μέχρι στιγμής αλγόριθμοι παραγωγής τέτοιων μοντέλων που είδαμε και παραπάνω συνήθως λαμβάνουν υπόψη έναν ή και δύο διαστάσεις το πολύ. Επομένως, χρήζει εύλογη η δημιουργία νέων αλγορίθμων που να στοχεύει σε μια τέτοια ισορροπία μεταξύ των διαστάσεων ποιότητας. [21]

2.2.5 Έλεγχος συμμόρφωσης.

Τα τελευταία χρόνια έχουν αναπτυχθεί ισχυρές τεχνικές εξόρυξης διαδικασιών που μπορούν και κατασκευάζουν αυτόματα ένα επιχειρησιακό μοντέλο διεργασιών, δεδομένου κάποιου αρχείου καταγραφής γεγονότων. Ενώ είδαμε ότι η ανακάλυψη διαδικασιών κατασκευάζει ένα τέτοιο μοντέλο, η τεχνική ελέγχου συμμόρφωσης χρησιμοποιεί το παραγόμενο μοντέλο και το αντίστοιχο αρχείο καταγραφής γεγονότων (που χρησιμοποιήθηκε ως είσοδο για τη δημιουργία του μοντέλου). Έτσι, συγκρίνεται η μοντελοποιημένη συμπεριφορά μαζί με την παρατηρούμενη συμπεριφορά του αρχείου καταγραφής και παίρνουμε ως αποτέλεσμα κάποιες διαγνωστικές πληροφορίες που μας δείχνουν τις διαφορές αλλά και τις ομοιότητες μεταξύ αυτών των δύο. Η τεχνική του ελέγχου συμμόρφωσης καταφέρνει και αναγνωρίζει την ικανότητα αντιστοιχίας του επιχειρησιακού μοντέλου με το αρχείο καταγραφής γεγονότων. [24]

2.2.5.1 Περιπτώσεις εφαρμογής ελέγχου συμμόρφωσης

Οι τεχνικές ελέγχου συμμόρφωσης συσχετίζουν συμβάντα του αρχείου καταγραφής με δραστηριότητες στο επιχειρησιακό μοντέλο. Με αυτόν τον τρόπο, πραγματοποιείται η σύγκριση της παρατηρούμενης με της μοντελοποιημένης συμπεριφοράς. Έτσι, μπορεί κανείς να ποσοτικοποιήσει τις διαφορές που παρατηρούνται και να διαγνώσει αποκλίσεις μεταξύ μοντέλου και αρχείου καταγραφής γεγονότων (π.χ. στο αρχείο καταγραφής παραλείπεται συχνά μια δραστηριότητα ενώ το μοντέλο δεν το επιτρέπει). Μερικές περιπτώσεις που χρησιμοποιείται η τεχνική του ελέγχου συμμόρφωσης είναι:

1. Για τον έλεγχο της ποιότητας του παραγόμενου μοντέλου (πόσο καλά περιγράφει το αρχείο καταγραφής),
2. Για τον εντοπισμό αποκλίνουσων περιπτώσεων και την κατανόηση των κοινών στοιχείων,
3. Για τον εντοπισμό των διαδικασιών που εμφανίζουν αποκλίσεις,
4. Για ελέγχους,
5. Για την κρίση ποιότητας ενός μοντέλου,
6. Για τη βοήθεια βελτίωσης γενετικών αλγορίθμων ανακάλυψης μοντέλων διεργασιών,
7. Ως σημείο εκκίνησης της τεχνικής βελτίωσης μοντέλων.

[24]

2.2.5.2 Τεχνικές ελέγχου συμμόρφωσης.

Μέχρι στιγμής, υπάρχουν τρεις προσεγγίσεις για τον έλεγχο συμμόρφωσης στην εξόρυξη διαδικασιών, τις οποίες θα δούμε σε αυτό το κεφάλαιο. Στην πράξη, στο κεφάλαιο 4.3, θα δειχθούν μόνο οι δύο πιο διαδεδομένες τεχνικές συμμόρφωσης και θα παρατηρήσουμε αναλυτικά τον τρόπο εφαρμογής της κάθε μιας.

Η πρώτη προσέγγιση είναι γνωστή ως casual footprints. Ουσιαστικά, πρόκειται για έναν πίνακα ο οποίος περιέχει όλες τις σχέσεις εξάρτησης μεταξύ των δραστηριοτήτων (εικόνα 17). Μόλις συμπληρωθούν οι δύο πίνακες (ένας για το μοντέλο και ένας για το αρχείο καταγραφής), συγκρίνονται μεταξύ τους με στόχο την εύρεση διαφορών. Για παράδειγμα, το αποτύπωμα (footprint) του μοντέλου μπορεί να δείχνει ότι μια δραστηριότητα ‘a’ ακολουθείται μερικές από μια άλλη ‘b’ αλλά ποτέ το αντίθετο. Εάν και το μοντέλο του αρχείου καταγραφής δείχνει το ίδιο τότε λέμε ότι τα αποτυπώματα συμπίπτουν αλλιώς ότι διαφέρουν και ουσιαστικά βρίσκουμε τις διαφορές μεταξύ αρχείου καταγραφής και μοντέλου. Ένας τύπος με τον οποίο μπορούμε να υπολογίσουμε την ποιότητα του μοντέλου είναι ο ακόλουθος:

$$\text{footprint based conformance} = 1 - \frac{\delta}{\kappa},$$

όπου δ είναι ο αριθμός των διαφορών που παρατηρήθηκαν και κ είναι ο αριθμός των κελιών που υπάρχουν στον πίνακα. [24]

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

Εικόνα 17. Παράδειγμα αποτυπώματος.

Η δεύτερη προσέγγιση (token based replay), αναπαράγει την κάθε περίπτωση που συναντάμε στο αρχείο καταγραφής γεγονότων, πάνω στο επιχειρησιακό μοντέλο. Πρόκειται για μια «απλή» διαδικασία με την οποία μετράμε τις περιπτώσεις που μπορούν και αναπαράγονται πλήρως από το μοντέλο. Επειδή αυτή η προσέγγιση δεν μπορεί να αναγνωρίσει τις περιπτώσεις που αναπαράγονται μέχρι κάποιο βαθμό από αυτές που δεν αναπαράγονται καθόλου, η προσέγγιση αυτή εφάρμοσε τη τεχνική των μαρκών. Συνεχίζει η αναπαραγωγή της περίπτωσης πάνω στο μοντέλο ακόμα και στις περιπτώσεις που δεν μπορούν να συνεχίσουν να αναπαράγονται απλά «με τον δανεισμό μιας μάρκας» και καταγράφουμε το πρόβλημα. Στο τέλος, θα υπολογιστεί η καταλληλότητα (fitness) του μοντέλου με βάση τις μάρκες αυτές. Περισσότερα για τη διαδικασία αυτή βρίσκει κανείς στο κεφάλαιο 4.3 [24]

Η τρίτη και τελευταία προσέγγιση (alignments), αποτελεί μια πιο «προχωρημένη» προσέγγιση σε σχέση με τις προηγούμενες και στόχος της είναι η δημιουργία μιας «ευθυγράμμισης» μεταξύ μιας περίπτωσης του αρχείου καταγραφής και του μοντέλου. Πιο συγκεκριμένα, η παρακάτω εικόνα (εικόνα 18) απεικονίζει τρία διαφορετικά παραδείγματα περιπτώσεων της τεχνικής αυτής. Η πάνω σειρά περιγράφει τις κινήσεις των διεργασιών μιας περίπτωσης του αρχείου καταγραφής γεγονότων ενώ η κάτω σειρά περιγράφει τις κινήσεις του μοντέλου.

$$\gamma_1 = \begin{array}{|c|c|c|c|c|} \hline a & c & d & e & h \\ \hline a & c & d & e & h \\ \hline \end{array} \quad \gamma_2 = \begin{array}{|c|c|c|c|c|} \hline a & d & c & e & h \\ \hline a & \gg & c & e & h \\ \hline \end{array} \quad \gamma_3 = \begin{array}{|c|c|c|c|c|c|c|c|c|} \hline a & d & c & e & f & d & b & e & h \\ \hline a & \gg & c & e & f_2 & \gg & b & e & h \\ \hline \end{array}$$

Εικόνα 18. Παραδείγματα τεχνικής alignments.

Στην περίπτωση γ_1 , έχουμε τέλεια ευθυγράμμιση διότι όλες οι κινήσεις της περίπτωσης μπορούν να γίνουν και από το μοντέλο. Στο γ_2 , παρατηρούμε ότι μόνο η κίνηση a,c,e,h μπορεί να παρατηρηθεί στο μοντέλο. Στο d, που δεν μπορεί να αποδοθεί από το μοντέλο, μπαίνει στη θέση του το σύμβολο '>>' και δηλώνει την κίνηση αυτή (απουσία δυνατότητας μίμησης της κίνησης από το μοντέλο). Αντίστοιχα στο γ_3 , με δύο αυτή τη φορά κινήσεις που δεν μπορούν να αποδοθούν από το μοντέλο (f=f2). [24]

Τη συμμόρφωση μπορεί κανείς να τη δει από δύο οπτικές γωνίες. Είτε το μοντέλο δεν αποτυπώνει την πραγματική συμπεριφορά και άρα «είναι λάθος», είτε η πραγματικότητα

αποκλίνει από το επιθυμητό μοντέλο άρα το αρχείο καταγραφής γεγονότων «είναι λάθος». Η πρώτη οπτική λαμβάνεται όταν το μοντέλο υποτίθεται ότι είναι περιγραφικό (δηλαδή καταγράφει την πραγματικότητα) και η δεύτερη οπτική γωνία λαμβάνεται υπόψη όταν το μοντέλο είναι κανονιστικό (δηλαδή ότι χρησιμοποιείται για να επηρεάσει ή να ελέγξει την πραγματικότητα). [24]

2.3 Μπεϋζιανά Δίκτυα (Bayesian Networks)

2.3.1 Εισαγωγή

Μερικές φορές πρέπει να υπολογιστεί η πιθανότητα μιας αβέβαιης αιτίας, δεδομένου κάποιων παρατηρούμενων στοιχείων. Για παράδειγμα, θέλουμε να γνωρίζουμε την πιθανότητα μιας συγκεκριμένης ασθένειας όταν παρατηρούνται συμπτώματα σε έναν ασθενή. Ένα τέτοιο πρόβλημα, συχνά είναι πολύπλοκο και με πολλές αλληλένδετες μεταβλητές, αφού μπορεί να υπάρχουν αρκετά συμπτώματα και ακόμα περισσότερες αιτίες. Στην πράξη, συνήθως είναι εύκολο να επιτευχθεί μόνο η υπό όρους αντίστροφη πιθανότητα, δηλαδή η πιθανότητα παρατήρησης συμπτωμάτων δεδομένης μιας αιτίας, ή η πιθανότητα παρατήρησης των συμπτωμάτων δεδομένου ότι ο ασθενής έχει μια συγκεκριμένη ασθένεια. Μια προσέγγιση χρησιμοποιώντας ένα μπεϋζιανό δίκτυο (Bayesian approach), είναι κατάλληλη για αυτές τις περιπτώσεις καθώς τα μπεϋζιανά δίκτυα είναι πολύ χρήσιμα εργαλεία για την αντιμετώπιση όχι μόνο της αβεβαιότητας αλλά και της πολυπλοκότητας και αιτιότητας. [26]

2.3.2 Θεωρία πιθανοτήτων

Πριν προχωρήσουμε στα μπεϋζιανά δίκτυα, είναι σκόπιμη μια μικρή εισαγωγή σε κάποιους βασικούς όρους που αφορούν τις πιθανότητες, οι οποίοι θα χρησιμοποιηθούν και στην πορεία στα μπεϋζιανά δίκτυα.

- (1) **Πείραμα τύχης:** πρόκειται για κάποιο πείραμα στο οποίο δεν μπορούμε εκ των προτέρων να προβλέψουμε το αποτέλεσμα, μολονότι επαναλαμβάνονται κάτω από τις ίδιες συνθήκες.
- (2) **Δειγματικός χώρος:** πρόκειται για το σύνολο των αποτελεσμάτων που μπορούν να εμφανιστούν σε ένα δειγματικό χώρο.
- (3) **Ενδεχόμενα:** πρόκειται για το σύνολο που έχει ως στοιχεία ένα ή περισσότερα αποτελέσματα ενός πειράματος τύχης.
- (4) **Πράξεις με ενδεχόμενα:** οι τέσσερις βασικές πράξεις που παρατηρούνται μεταξύ ενδεχομένων είναι:

1. $A \cap B$, “A τομή (και) B”, πραγματοποιούνται συγχρόνως τα A και B.
2. $A \cup B$, “A ένωση (ή) B” πραγματοποιείται ένα τουλάχιστον από τα A, B.
3. A' , “όχι A” ή “συμπληρωματικό του A”, ή “αντίθετο του A” και πραγματοποιείται όταν δεν πραγματοποιείται το A.
4. $A - B$, “διαφορά του B από το A” και πραγματοποιείται, όταν πραγματοποιείται το A αλλά όχι το B. Είναι εύκολο να δούμε ότι $A - B = A \cap B'$.

(5) Ασυμβίβαστα ενδεχόμενα: δύο ενδεχόμενα A και B λέγονται ασυμβίβαστα, όταν $A \cap B = 0$.

(6) Κλασικός ορισμός πιθανότητας: σε ένα πείραμα τύχης με n ισοπίθανα αποτελέσματα, η πιθανότητα ενός ενδεχομένου A που περιέχει k τέτοια αποτελέσματα είναι: $P(A) = \frac{\text{πλήθος εννοϊκών αποτελεσμάτων για το A}}{\text{πλήθος όλων των δυνατών αποτελεσμάτων}} = \frac{k}{n}$

Είναι ιδιαίτερα συνηθισμένο σε πειράματα τύχης τα αποτελέσματα να μην είναι ισοπίθανα. Σε τέτοιες περιπτώσεις δεν μπορεί να χρησιμοποιηθεί ο παραπάνω ορισμός πιθανότητας αλλά χρησιμοποιείται ο **(7) αξιωματικός ορισμός της πιθανότητας**, ο οποίος έχει ως εξής:

Έστω $\Omega = \{\omega_1, \omega_2, \dots, \omega_n\}$ ένας δειγματικός χώρος ενός πειράματος τύχης με πεπερασμένο πλήθος στοιχείων. Σε κάθε απλό ενδεχόμενο $\{\omega_i\}$ αντιστοιχίζουμε έναν πραγματικό αριθμό, που τον συμβολίζουμε με $P(\omega_i)$, έτσι ώστε να ισχύουν:

- $0 \leq P(\omega_i) \leq 1$
- $P(\omega_1) + P(\omega_2) + \dots + P(\omega_n) = 1$

Τον αριθμό $P(\omega_i)$ ονομάζουμε πιθανότητα του ενδεχομένου $\{\omega_i\}$. Ως πιθανότητα $P(A)$ ενός ενδεχομένου $A = \{\alpha_1, \alpha_2, \dots, \alpha_k\} \neq 0$, ορίζουμε το άθροισμα $P(\alpha_1) + P(\alpha_2) + \dots + P(\alpha_k)$, ενώ ως πιθανότητα του αδύνατου ενδεχομένου 0 , ορίζουμε τον αριθμό $P(0) = 0$.

Στην περίπτωση που το $P(\omega_i) = \frac{1}{n}, i = 1, 2, \dots, n$, τότε λέμε ότι έχουμε τον κλασικό ορισμό πιθανότητας ενός ενδεχομένου.

(8) Δεσμευμένη πιθανότητα: πρόκειται για την πιθανότητα εμφάνισης ενός ενδεχομένου A δεδομένου ότι έχει ή (ή θα) πραγματοποιηθεί ένα άλλο ενδεχόμενο B (η πραγματοποίηση του B επηρεάζει την πιθανότητα πραγματοποίησης του A). Συμβολίζεται

με $A | B$ και η πιθανότητα $P(A|B) = \frac{P(A \cap B)}{P(B)}$ αντίστοιχα. Άμεση συνέπεια του ορισμού αυτού είναι: $P(A \cap B) = P(A) * P(B|A) = P(B) * P(A|B)$.

(9) Ανεξάρτητα ενδεχόμενα: πρόκειται για τα ενδεχόμενα που η πληροφορία για την πραγματοποίηση του ενός δεν επηρεάζει την πιθανότητα πραγματοποίησης του άλλου. $P(A|B) = P(A)$ και $P(B|A) = P(B)$.

[27]

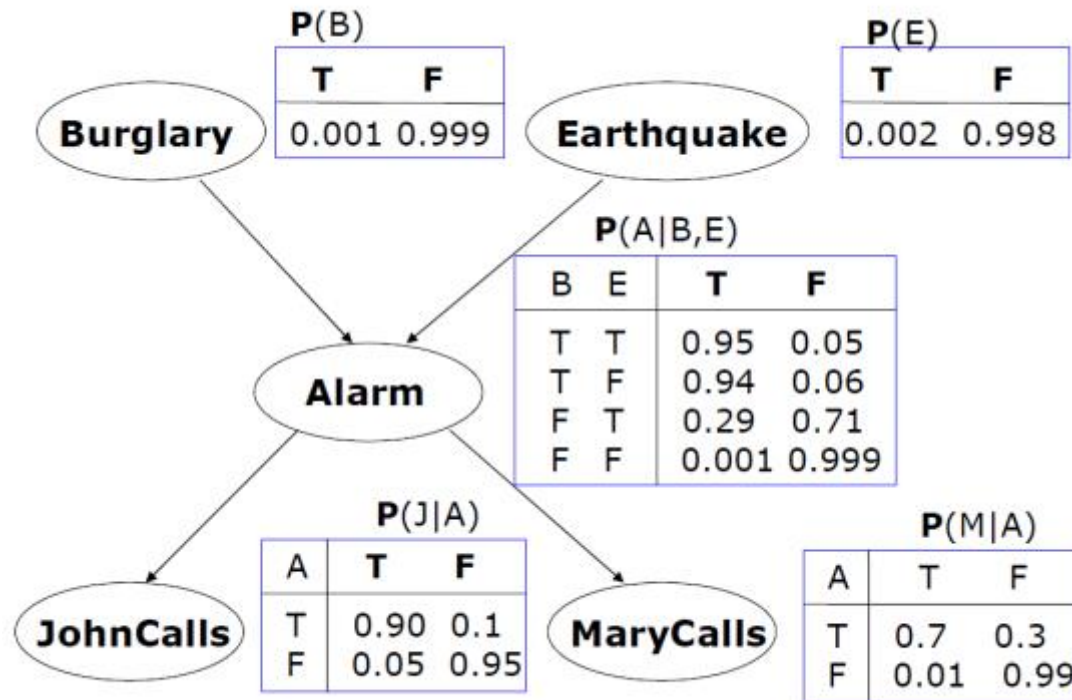
2.3.3 Θεωρητικό υπόβαθρο Bayesian Networks

Το μπεϋζιανό δίκτυο (Bayesian network ή BN), γνωστό και ως δίκτυο πεποίθησης ανήκει στην οικογένεια των πιθανολογικών γραφικών μοντέλων και αντιστοιχίζεται με μια ιδιαίτερη δομή, γνωστή και ως κατευθυνόμενο άκυκλο γράφημα (DAG). Τα γραφικά μοντέλα χρησιμοποιούνται για να την αναπαράσταση γνώσης σχετικά με ένα αβέβαιο τομέα. Πιο συγκεκριμένα, κάθε κόμβος σε ένα τέτοιο γράφημα αντιπροσωπεύει μια τυχαία μεταβλητή ενώ οι ακμές μεταξύ κόμβων αντιπροσωπεύουν τις πιθανολογικές εξαρτήσεις μεταξύ των κόμβων αυτών. Αυτές οι υπό όρους εξαρτήσεις στο γράφημα υπολογίζονται συχνά χρησιμοποιώντας γνωστές στατιστικές και υπολογιστικές μεθόδους. Ως εκ τούτου, τα μπεϋζιανά δίκτυα συνδυάζουν αρχές από τη θεωρία γραφημάτων, τη θεωρία πιθανότητας, την επιστήμη των υπολογιστών και τη στατιστική. [28]

Εκτός από τη δομή του μπεϋζιανού δικτύου (DAG), που θεωρείται το «ποιοτικό» μέρος του μοντέλου, χρειάζεται να προσδιοριστούν και οι «ποσοτικοί» παράμετροι του μοντέλου. Οι παράμετροι αυτοί περιγράφονται από την ιδιότητα Μαρκόβ (Markovian property), δηλαδή η κατανομή των υπό όρων πιθανοτήτων σε κάθε κόμβο εξαρτάται αποκλειστικά και μόνο από τους «γονείς» κόμβους. Συνήθως, η υπό όρους πιθανότητα αντιπροσωπεύεται από έναν πίνακα (Conditional Probability Table – CPT), στον οποίο αναγράφεται η πιθανότητα εμφάνισης ενός κόμβου «παιδί» ανάλογα τον συνδυασμό εμφάνισης των «γονικών» του κόμβων. [28]

Ένα μπεϋζιανό δίκτυο αντικατοπτρίζει μια απλή δήλωση ανεξαρτησίας υπό όρους. Δηλαδή, κάθε μεταβλητή είναι ανεξάρτητη από τους μη απογόνους της στο γράφημα, δεδομένης της κατάστασης των γονικών της κόμβων. Αυτή η ιδιότητα χρησιμοποιείται για τη μείωση, μερικές φορές, του αριθμού των παραμέτρων που απαιτούνται για τον

χαρακτηρισμό της πιθανότητας εμφάνισης δύο ή παραπάνω μεταβλητών ταυτόχρονα. Αυτή η μείωση παρέχει έναν αποτελεσματικό τρόπο υπολογισμού των οπίσθιων πιθανοτήτων, λαμβανομένων υπόψη των στοιχείων. [28]



Εικόνα 19. Παράδειγμα δομής μπεϋζιανού δικτύου. [28]

Με βάση τα παραπάνω, ένας πιο επιστημονικός ορισμός θα μπορούσε να έχει ως εξής: ένα μπεϋζιανό δίκτυο είναι ένα άκυκλο γράφημα που αντιπροσωπεύει την κατανομή των πιθανοτήτων εμφάνισης δύο ή και παραπάνω τυχαίων μεταβλητών V ταυτόχρονα (joint probability distribution). Το δίκτυο ορίζεται από ένα ζεύγος $B = \{ G, \Theta \}$, όπου το G είναι το άκυκλο κατευθυνόμενο γράφημα (DAG) του οποίου οι κόμβοι X_1, X_2, \dots, X_n αντιπροσωπεύουν τυχαίες μεταβλητές και οι ακμές τους που αντιπροσωπεύουν τις άμεσες εξαρτήσεις μεταξύ των μεταβλητών αυτών. Το γράφημα G κωδικοποιεί παραδοχές ανεξαρτησίας, με τις οποίες κάθε μεταβλητή X_i είναι ανεξάρτητη από τους μη απογόνους του, δεδομένου των γονέων του. Το Θ δηλώνει το σύνολο των παραμέτρων του δικτύου. Τέλος, ο τύπος υπολογισμού της πιθανότητας πραγματοποίησης όλων των μεταβλητών του B , γράφεται ως εξής:

$$P(X_1, X_2, \dots, X_n) = \prod_{i=1}^n P(X_i | \text{parent}(X_i)).$$

[28]

2.3.4 Θεώρημα Bayes

Το θεώρημα Μπέυζ (Bayes' Theorem ή και Bayes' Rule) είναι ένα από τα γνωστότερα θεωρήματα στο χώρο της στατιστικής και των μαθηματικών και διατυπώθηκε από τον Thomas Bayes (1701-1761). Έστω ότι έχουμε δύο διαφορετικά γεγονότα, A και B και γνωρίζουμε την υπό όρους πιθανότητα εμφάνισης του A δεδομένου του B, γραμμένο και ως $P(A|B)$. Χρησιμοποιούμε το θεώρημα του Μπέυζ για να βρούμε το $P(B|A)$, την υπό όρους πιθανότητα εμφάνισης του B, δεδομένου του A. Έτσι:

$$\text{Bayes' Theorem: } P(B|A) = \frac{P(A|B) * P(B)}{P(A)}$$

Όπου $P(A)$ και $P(B)$ οι πιθανότητες εμφάνισης του A και B αντίστοιχα.

Μια επίσης, διαφορετική μορφή, του θεωρήματος αυτού που μπορεί να βρεθεί είναι η:

$$P(B|A) = \frac{P(A|B) * P(B)}{P(A|B) * P(B) + P(A|B') * P(B')},$$

όπου $P(B')$ η πιθανότητα του B να μην πραγματοποιηθεί.

Μια ακόμα μορφή που μπορεί να πάρει το θεώρημα Μπέυζ στην περίπτωση τριών διαφορετικών γεγονότων είναι η ακόλουθη:

$$P(A|B, C) = \frac{P(B|A, C) * P(A|C)}{P(B|C)}.$$

Ένα απλό παράδειγμα για την κατανόηση του παραπάνω τύπου θα μπορούσε να είναι το εξής πρόβλημα:

Να βρεθεί η πιθανότητα εμφάνισης φωτιάς δεδομένου ότι υπάρχει καπνός, όταν η πιθανότητα εμφάνισης φωτιάς είναι 1%, η πιθανότητα εμφάνισης καπνού είναι 10% και γνωρίζουμε ότι στο 90% των περιπτώσεων φωτιάς παρατηρούμε καπνό.

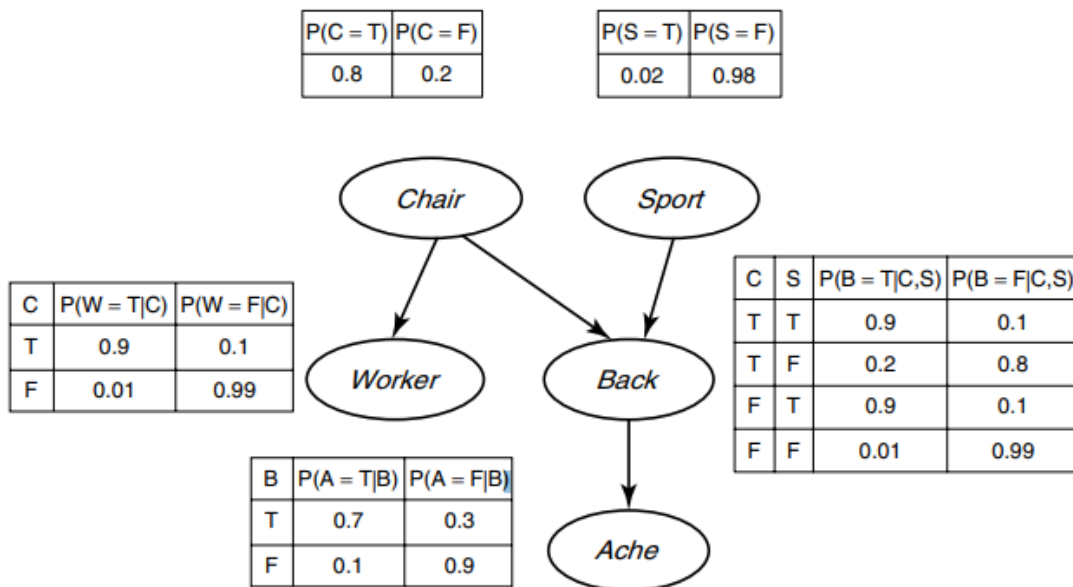
Χρησιμοποιώντας το θεώρημα Μπέυζ έχουμε:

$$P(\text{φωτιά}|\text{καπνός}) = \frac{P(\text{καπνός}|\text{φωτιά}) * P(\text{φωτιά})}{P(\text{καπνός})} = \frac{90\% * 1\%}{10\%} = 9.$$

[29, 30]

2.3.5 Συμπερασμός (Inference)

Μόλις ολοκληρωθεί το μπεϋζιανό δίκτυο, μπορεί να χρησιμοποιηθεί για να απαντήσει διάφορες «ερωτήσεις», σχετικά με τις πιθανότητες εμφάνισης κάποιας μεταβλητής που ενδιαφερόμαστε (πιθανότητες ενδιαφέροντος). Για παράδειγμα, στο πρόβλημα που απεικονίζεται στην εικόνα 19, θέλουμε να υπολογίσουμε την πιθανότητα να έχει γίνει διάρρηξη, δεδομένου παρατήρησης πραγματοποίησης άλλων μεταβλητών. Μια τέτοια πιθανότητα δεν βρίσκεται αποθηκευμένη στο μπεϋζιανό δίκτυο που δημιουργήθηκε, αλλά πρέπει να υπολογιστεί. Η διαδικασία του υπολογισμού μιας πιθανότητας ενδιαφέροντος, δεδομένου γνωστών αλλά και νέων στοιχείων, καλείται συμπερασμός. [32, 33]



Εικόνα 20. Παράδειγμα BN για συμπερασμό (chair = c, sport = s, worker = w, back = b, ache = a). [32]

Στο παράδειγμα της παραπάνω εικόνας (εικόνα 20), αν θέλαμε να υπολογίσουμε το $P(c=T|a=T)$, τότε θα λέγαμε:

$$P(c = T|a = T) = \frac{P(c = T, a = T)}{P(a = T)}$$

όπου

$$P(c = T, a = T) = \sum_{s,w,b} P(c = T) * P(s) \times P(w|c = T) * P(b|s, c = T) * P(a = T|b)$$

και

$$P(a = T) = \sum_{s,w,b,c} P(c) * P(s) * P(w|c) * P(b|s, c) \times P(a = T|b)$$

[32]

Δεδομένου ότι ο παραπάνω υπολογισμός είναι εκθετικά πολύπλοκος, έχουν δημιουργηθεί αλγόριθμοι εξαγωγής συμπερασμού. Η πιο γνωστή μέθοδος είναι η απαλοιφή μεταβλητών (Variable Elimination).

Συνοπτικά, η βασική ιδέα της μεθόδου αυτής είναι να αφαιρεί μία μία τις μεταβλητές που δεν χρειάζονται. Κάθε φορά που αφαιρείται μια μεταβλητή, δημιουργείται μια νέα σχέση που αντικατοπτρίζει τη σχέση της μεταβλητής με τις υπόλοιπες, μειώνοντας έτσι και το μέγεθος. Επίσης, οι πράξεις μένουν στη μνήμη και κάθε φορά που συναντάει ο αλγόριθμος μια «γνωστή» πράξη, δεν την επαναυπολογίζει. Τέλος, μειώνοντας τις σχέσεις που πρέπει να υπολογιστούν, μειώνεται και η υπολογιστική ισχύ που χρειαζόμαστε για τους υπολογισμούς. [32]

3. Προτεινόμενη προσέγγιση

Οι επιχειρησιακές διαδικασίες αποτελούν το πιο σημαντικό στοιχείο μιας επιχείρησης, καθώς καθορίζουν μια δραστηριότητα ή ένα σύνολο δραστηριοτήτων που έχει σκοπό η επιχείρηση να ολοκληρώσει με επιτυχία, ως οργανωτικό στόχο. Η μοντελοποίηση και η ανάλυση των διαδικασιών είναι κρίσιμη για τον προσδιορισμό των υπάρχουσών διαδικασιών και την κατανόηση των συνεισφορών των νέων διαδικασιών στο σύστημα. [10]

Οι τεχνικές εξόρυξης διεργασιών χρησιμοποιούνται για την ανακάλυψη και την ανάλυση των επιχειρησιακών διαδικασιών με αυτοματοποιημένο τρόπο. Αφού προσπαθήσουν να ανακαλύψουν τη δομή, οργάνωση και ιδιότητες της επιχειρησιακής διαδικασίας μπορούν να απεικονιστούν ως μοντέλα επιχειρησιακών διαδικασιών. [10]

Πολλές όμως είναι οι φορές που κάποιο παραγόμενο μοντέλο δεν «ταυτίζεται» με τις υπάρχουσες διαδικασίες της επιχείρησης καθώς, είτε υπάρχουν σχέσεις μεταξύ δραστηριοτήτων που φαινομενικά δεν μπορούν να πραγματοποιηθούν είτε δεν υπάρχουν στο σύστημα της επιχείρησης. Και όλα αυτά γιατί από μόνη της μια διαδικασία μπορεί να γίνει αρκετά πολύπλοκη από τη φύση της, όταν προσπαθήσουμε να την εφαρμόσουμε σε κάποιο σύστημα.

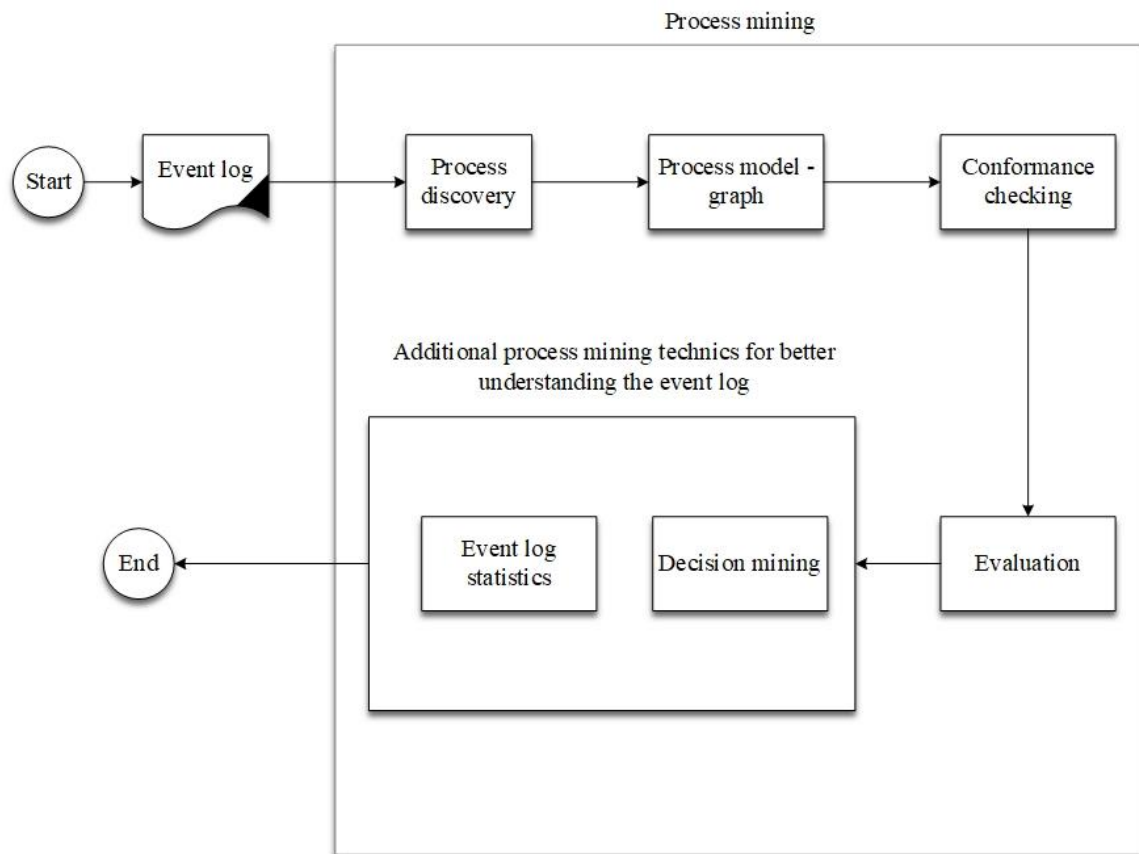
Για το χειρισμό των παραπάνω προβλημάτων αλλά και γενικότερα της αβεβαιότητας που συναντάμε στις επιχειρησιακές διαδικασίες, μπορούμε να χρησιμοποιήσουμε κάποιο πιθανοτικό μοντέλο μάθησης. Στην παρούσα διπλωματική, χρησιμοποιήθηκαν τα μπεϋζιανά δίκτυα.

Τα μπεϋζιανά δίκτυα μπορούν να φανούν αρκετά χρήσιμα στον τομέα της εξόρυξης διαδικασιών. Ένα από τα μεγαλύτερα πλεονεκτήματα της εφαρμογής τους, είναι στις περιπτώσεις που δεν μπορούμε να γνωρίζουμε εάν έχει πραγματοποιηθεί μια διεργασία και θέλουμε να μάθουμε την πιθανότητα της επιχειρησιακής διαδικασίας να τερματιστεί ή να φτάσει σε μια επόμενη διεργασία. Δηλαδή σε περιπτώσεις που κυριαρχεί η αβεβαιότητα. Μπορούν να αντιπροσωπεύσουν τις επιχειρησιακές διαδικασίες ως κόμβους (μεταβλητές) και τις μεταβάσεις/σειρά των διαδικασιών ως τις ακμές μεταξύ τους. Οι πίνακες των πιθανοτήτων υπό όρους μπορούν αυτόματα να υπολογιστούν από το αρχείο καταγραφής γεγονότων καθώς και με τη βοήθεια κάποιου μπεϋζιανού ταξινομητή. [34]

Οι μέχρι στιγμής οι υπάρχοντες αλγόριθμοι εξόρυξης διεργασιών δεν μπορούν να δημιουργήσουν κάποιο μοντέλο επιχειρησιακών διεργασιών που να μπορεί να χρησιμοποιηθεί ως είσοδος σε κάποιο μπεϋζιανό δίκτυο, λόγω ότι ένα δίκτυο πεποίθησης ή γνωστό σε όλους μπεϋζιανό δίκτυο, πρόκειται για άκυκλο γράφημα. [10]

Επομένως, μπορούμε να χωρίσουμε την προσέγγιση σε δύο μέρη. Ως πρώτο μέρος θα ορίσουμε την εξόρυξη διεργασιών και ουσιαστικά θα εφαρμοστούν τεχνικές εξόρυξης με στόχο την ανακάλυψη και ανάλυση των διεργασιών. Ως δεύτερο μέρος, θα ορίσουμε τη μετατροπή κάποιου επιχειρησιακού μοντέλου, που έχουμε πάρει από το πρώτο μέρος της προσέγγισης, σε ένα μπεϋζιανό δίκτυο και με απώτερο στόχο προβλέψουμε την εμφάνιση κάποιας δραστηριότητας.

Ο αλγόριθμος που ακολουθήθηκε για το πρώτο μέρος της προσέγγισης φαίνεται στην εικόνα 21.



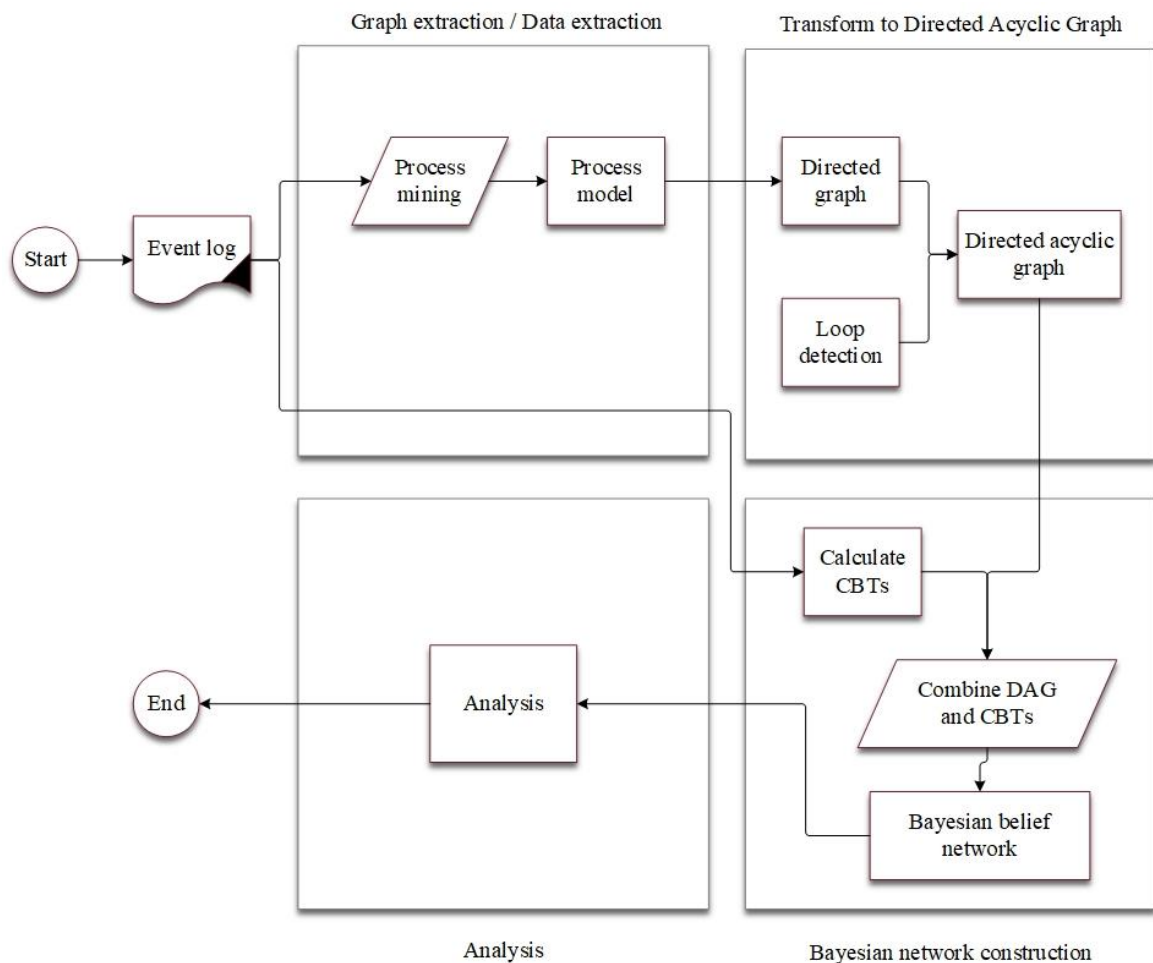
Εικόνα 21. Αλγόριθμος για εξαγωγή μοντέλου διεργασιών και περισσότερα σε εξόρυξη διεργασιών.

Ουσιαστικά έχουμε τέσσερα βήματα, δίνοντας μεγαλύτερη έμφαση στο πρώτο, αφού πρόκειται να χρησιμοποιηθεί και στο δεύτερο μέρος της προσέγγισης. Ξεκινώντας από ένα αρχείο καταγραφής γεγονότων, ακολουθούμε τα παρακάτω βήματα:

1. Μέσω των αλγορίθμων ανακάλυψης διεργασιών, εξάγουμε ένα κατάλληλο μοντέλο διεργασιών, που θα περιγράφει τη σειρά των δραστηριοτήτων που εκτελούνται κατά τη διάρκεια εκτέλεσης μιας διεργασίας.
2. Μέσω ενός συνόλου τεχνικών συμμόρφωσης, συγκρίνουμε τα παραγόμενα μοντέλα διεργασιών με το αρχείο καταγραφής γεγονότων. Ο στόχος μας είναι να ελέγξουμε εάν το αρχείο καταγραφής γεγονότων συμμορφώνεται με το επιχειρησιακό μοντέλο και αντιστρόφως.
3. Συγκρίνουμε τη συμπεριφορά που παρατηρούμε στο αρχείο καταγραφής γεγονότων και στο επιχειρησιακό μοντέλο για να δούμε κατά πόσο και αν ταιριάζουν. Τέσσερις είναι οι βασικές «διαστάσεις» που μετρούνται, η ικανότητα της επανάληψης, η ακρίβεια, η γενίκευση και η μέτρηση της απλότητας.
4. Επιπρόσθετες τεχνικές εξόρυξης διεργασιών για περεταίρω μελέτη του αρχείου καταγραφής γεγονότων.

Ο απώτερος στόχος της συγκεκριμένης διπλωματικής εργασίας είναι η δημιουργία ενός μπεϋζιανού δικτύου από κάποιο αρχείο καταγραφής γεγονότων, μιας επιχείρησης, ώστε να προβλέψουμε κάποια δραστηριότητα.

Όπως κάθε μπεϋζιανό δίκτυο αποτελείται από κάποιο κατευθυνόμενο άκυκλο γράφημα, έτσι και η προσέγγιση που ακολουθήθηκε στο δεύτερο μέρος εστιάζει στην εξαγωγή γραφήματος, από αρχείο καταγραφής γεγονότων, ύστερα από την αφαίρεση τυχόν κύκλων. Ο αλγόριθμος αυτός φαίνεται παρακάτω στην εικόνα 22.



Εικόνα 22. Αλγόριθμος για μετατροπή αρχείου καταγραφής συμβάντων σε Bayesian Network.

Το δεύτερο και κυριότερο μέρος της προσέγγισης ξεκινάει από ένα αρχείο καταγραφής γεγονότων (event log) που εξήχθη από το πληροφοριακό σύστημα μιας επιχείρησης. Αποτελείται από τέσσερα κύρια στάδια: (i) εξαγωγή γραφήματος και δεδομένων, (ii) μετατροπή γραφήματος σε κατευθυνόμενο άκυκλο γράφημα, (iii) κατασκευή μπεϋζιανού δικτύου, (iv) ανάλυση.

Πιο αναλυτικά, έχουμε τα εξής βήματα:

- Στάδιο 1, εξαγωγή γραφήματος και δεδομένων:
 1. Μέσω τεχνικών εξόρυξης διεργασιών εξάγουμε κάποιο μοντέλο διεργασιών.
- Στάδιο 2, μετατροπή γραφήματος σε κατευθυνόμενο άκυκλο γράφημα:
 2. Χρησιμοποιούμε το παραπάνω μοντέλο διεργασιών ως κατευθυνόμενο γράφημα.

3. Αναγνωρίζονται τυχόν κύκλοι μεταξύ των δραστηριοτήτων στο παραπάνω γράφημα και αφαιρούνται.
 4. Παράγουμε το κατευθυνόμενο άκυκλο γράφημα.
- Στάδιο 3, κατασκευή μπεϋζιανού δικτύου:
 5. Τα δεδομένα του αρχείου καταγραφής συμβάντων μετατρέπονται σε πίνακες δεσμευμένων πιθανοτήτων (Conditional probability tables-CBT).
 6. Γίνεται συνδυασμός των πινάκων δεσμευμένων πιθανοτήτων (Conditional probability tables-CBT) με του κατευθυνόμενου άκυκλου γραφήματος.
 7. Παράγουμε το μπεϋζιανό δίκτυο.
 - Στάδιο 4, ανάλυση:
 8. Εξαγωγή συμπερασμάτων και προβλέψεων.

3.1 Μέρος Α

3.1.1 Διαδικασία ανακάλυψης επιχειρησιακού μοντέλου.

Το πρώτο βήμα του Α' μέρους της προσέγγισής μας είναι η εξαγωγή κάποιου γραφήματος επιχειρησιακού μοντέλου διεργασιών από ένα αρχείο καταγραφής συμβάντων. Το συγκεκριμένο βήμα μας είναι ιδιαίτερα σημαντικό καθώς, ένα τέτοιο μοντέλο θα χρειαστούμε και παρακάτω για την κατασκευή του μπεϋζιανού δικτύου.

Υπάρχουν αρκετοί αλγόριθμοι εξαγωγής κάποιου επιχειρησιακού μοντέλου διεργασιών σε μορφή γραφήματος Petri net, τους οποίους θα δούμε αναλυτικά στο κεφάλαιο 4. Οι τρεις κύριοι αλγόριθμοι ανακάλυψης μοντέλων διεργασιών με τους οποίους θα ασχοληθούμε είναι οι: Alpha, Heuristic και Inductive miner. Επίσης, ένα επιπλέον γράφημα μοντέλου διεργασιών που θα δειχθεί είναι το Directly-Follows Graph. Συνολικά θα δειχτούν 4 γραφήματα σε μορφή Petri net, από τα οποία θα διαλέξουμε ένα από αυτά (αυτό που μας βολεύει καλύτερα για τη περίπτωση μας) για να συνεχίσουμε στο Β' μέρος της προσέγγισης.

3.1.2 Έλεγχος συμμόρφωσης (Conformance checking)

Σε αυτό το βήμα, συγκρίνουμε τα παραγόμενα μοντέλα διεργασιών με το αρχείο καταγραφής γεγονότων. Ο στόχος μας είναι να ελέγξουμε εάν το αρχείο καταγραφής γεγονότων συμμορφώνεται με το επιχειρησιακό μοντέλο και αντιστρόφως. Το αρχείο καταγραφής συμβάντων αναπαράγεται πάνω στο επιχειρησιακό μοντέλο για να βρεθούν αποκλίσεις και έτσι μετράμε τις διαφορετικές συμπεριφορές που παρατηρούνται στο αρχείο καταγραφής και στο μοντέλο. Επιπρόσθετα, μέσω των τεχνικών συμμόρφωσης μπορούμε να μετρήσουμε και την απόδοση των επιχειρησιακών μοντέλων στην αναπαράσταση των διεργασιών και να επιδιορθώσουμε το μοντέλο ώστε να ανταποκρίνεται καλύτερα στην πραγματικότητα του αρχείου καταγραφής συμβάντων. [9] Δύο είναι οι πιο σημαντικές τεχνικές ελέγχου συμμόρφωσης, τις οποίες και θα δούμε: token-based replay και alignments.

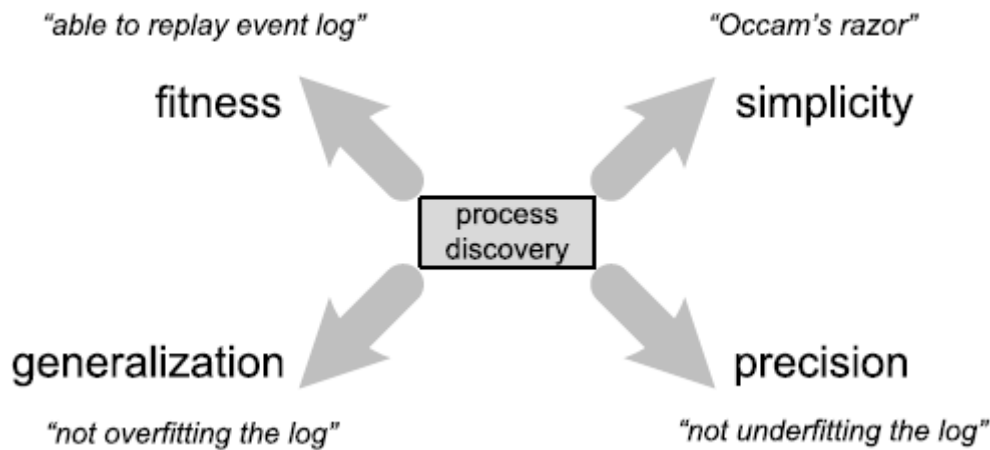
Και στις δύο περιπτώσεις θα ποσοτικοποιηθεί η έννοια της προσαρμογής (fitness), μιας και είναι ποιοτικό κριτήριο που σχετίζεται περισσότερο με τη συμμόρφωση. [9]

3.1.3 Αξιολόγηση επιχειρησιακών μοντέλων (Evaluation)

Στο παρόν βήμα, συγκρίνουμε τη συμπεριφορά που παρατηρούμε στο αρχείο καταγραφής γεγονότων και στο επιχειρησιακό μοντέλο για να δούμε κατά πόσο και αν ταιριάζουν. Τα τέσσερα κριτήρια που θα συγκριθούν είναι τα : fitness, simplicity, generalization και precision. Συνοπτικά θα μπορούσαμε να πούμε πως σε αυτό το βήμα ελέγχουμε το ποσοστό στο οποίο το παραγόμενο μοντέλο μας:

1. μπορεί να αναπαράξει όλες τις περιπτώσεις του αρχείο καταγραφής συμβάντων,
2. είναι ευανάγνωστο,
3. μπορεί να αναπαράξει μελλοντική συμπεριφορά κάποιας νέας περίπτωσης και
4. το ποσοστό που μας εγγυάται ότι δεν θα επιτρέψει νέα συμπεριφορά παραπάνω από ότι φαίνεται στο αρχείο καταγραφής συμβάντων. [9]

Μια σύνοψη των παραπάνω βλέπουμε στην εικόνα 23.



Εικόνα 23. Τα τέσσερα ανταγωνιστικά ποιοτικά κριτήρια. [9]

3.1.4 Περαιτέρω αναγνώσεις πληροφοριών αρχείων καταγραφής συμβάντων

Τα αρχεία καταγραφής συμβάντων, πέρα από τα αντικείμενα που έχουμε εξετάσει στο πρώτο μισό της προσέγγισης, περιέχουν επιπλέον μια πληθώρα πληροφοριών και θα ήταν κρίμα να μην παρατηρούσαμε και μια διαφορετική προοπτική, όπως την οργανωτική προοπτική ή την προοπτική του χρόνου ή και της περίπτωσης, μιας και μέχρι στιγμής «δουλεύαμε» κοιτάζοντας την προοπτική της ροής ελέγχου των διαδικασιών. [9]

Έτσι λοιπόν, το παρόν βήμα εξετάζει λοιπές προοπτικές, παρατηρώντας για παράδειγμα τη χρονική συχνότητα των δραστηριοτήτων ή και πως ο συνδυασμός των προοπτικών μπορεί να συγχωνευτεί σε ένα νέο μοντέλο έχοντας άποψη για μια διαδικασία. Ένα τέτοιο μοντέλο θα μπορούσε να χρησιμοποιηθεί για προσομοιώσεις με στόχο εξόρυξη αποφάσεων (decision mining). [9]

3.2 Μέρος Β

3.2.1 Εξαγωγή άκυκλου κατευθυνόμενου γραφήματος

Στο παρόν βήμα θα εξάγουμε ένα κατευθυνόμενο άκυκλο γράφημα και αυτό γιατί ένα μπεύζιανό δίκτυο εξ ορισμού αποτελεί ένα κατευθυνόμενο άκυκλο γράφημα.

Ήδη από το πρώτο μέρος της προσέγγισης έχουμε εξαγάγει, μέσω τεχνικών ανακάλυψης μοντέλων διαδικασιών, ορισμένα μοντέλα διαδικασιών. Τα μοντέλα αυτά, όπως είδαμε και από τη βιβλιογραφική επισκόπηση στο κεφάλαιο 2, αποτελούν ουσιαστικά κάποιο κατευθυνόμενο γράφημα, σε μορφή δικτύου Petri, που περιγράφει τη σειρά των δραστηριοτήτων που μπορούν και εκτελούνται κατά την εκτέλεση μιας διαδικασίας.

Επομένως, μπορούμε να διαλέξουμε ένα από τα παραπάνω γραφήματα δικτύου Petri και να προχωρήσουμε στη δημιουργία του μπεϋζιανού δικτύου μιας και λύνουμε έτσι εύκολα το θέμα του γραφήματος.

Πολλές φορές όμως, παίρνοντας αυτούσιο ένα τέτοιο επιχειρησιακό μοντέλο δεν είναι κατάλληλο, καθώς αν και είναι κατευθυνόμενο, το πρώτο κριτήριο που θέλουμε για το μπεϋζιανό δίκτυο, δεν είναι άκυκλο. Και αυτό γιατί τέτοια γραφήματα εμφανίζουν βρόγχους επανάληψης, μιας και τα αρχεία καταγραφής συμβάντων μπορεί να περιέχουν επαναλαμβανόμενες δραστηριότητες ή παράλληλες δραστηριότητες που να εμφανίζονται ως εξαρτώμενες στο μοντέλο ενώ στην πραγματικότητα να είναι ανεξάρτητες. [10]

Επιλέγοντας ένα γράφημα που έχουμε εξάγει από το πρώτο μέρος της προσέγγισης και σε συνδυασμό με το επόμενο ακολουθιακά βήμα της προσέγγισης (αφαίρεση κύκλων) παράγουμε ένα άκυκλο κατευθυνόμενο γράφημα που μπορούμε να χρησιμοποιήσουμε ως βάση στο μπεϋζιανό δίκτυο.

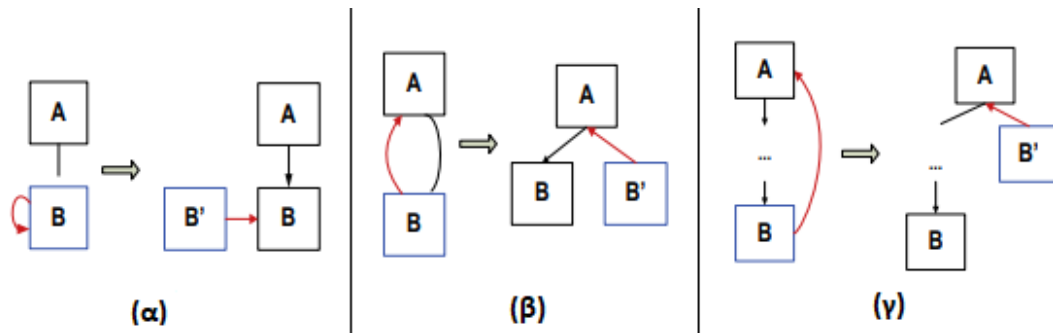
3.2.2 Αφαίρεση κύκλων

Σε αυτό το βήμα ασχολούμαστε με την αποσύνθεση βρόγχων που τυχόν υπάρχουν στο γράφημα.

Η γενική ιδέα που ακολουθήθηκε είναι η εξής: Για την αφαίρεση τυχόν κύκλων μεταξύ των δραστηριοτήτων που εμφανίζονται στο γράφημα, θα δημιουργείται ένα αντίγραφο της ουράς του κόμβου-δραστηριότητας στον βρόγχο. [12]

Ας το δούμε με ένα παράδειγμα. Παρακάτω στην εικόνα 24, βλέπουμε τρεις περιπτώσεις κύκλου, μεγέθους 1, 2 και n αντίστοιχα (5.α, 5.β, 5.γ). Στην πρώτη περίπτωση έχουμε την ακολουθία A,B,B,.. όπου το B εμφανίζεται αναδρομικά, στην δεύτερη περίπτωση συναντάμε την ακολουθία A,B,A,B,... όπου τα A και B εμφανίζονται εναλλάξ, ενώ στην

τελευταία πρόκειται για την περίπτωση όπου το A θα ξαναεμφανιστεί έπειτα από το B, ανεξαρτήτως ενδιάμεσων κόμβων. Για κάθε μια από τις περιπτώσεις βλέπουμε αντίστοιχα πως η δημιουργία ενός «όμοιου» κόμβου σπάει τον κύκλο χωρίς να χάνει τις πληροφορίες του. [12]



Εικόνα 24. Αποσύνθεση κύκλων μεγέθους 1, 2, n, αντίστοιχα. [12]

Προτιμήθηκε η τεχνική αυτή, δηλαδή αντιπροσώπευση ουσιαστικά κύκλου με πολλές εμφανίσεις του ίδιου κόμβου, έναντι άλλων λόγω ότι κρατάμε περισσότερα δεδομένα, περισσότερη πληροφορία από το αρχείο καταγραφής συμβάντων. Θα μπορούσαμε για παράδειγμα να εφαρμόζαμε κάποιον ευρετικό αλγόριθμο και σε περίπτωση κύκλου να κράταγε την ακμή με τις περισσότερες μεταβάσεις μεταξύ των κόμβων, μειώνοντας έτσι παραπάνω δεδομένα και πληροφορίες που παίρνουμε από το αρχείο καταγραφής συμβάντων. Δηλαδή εάν είχαμε την περίπτωση β της εικόνας 5 και λέγαμε ότι η διαδρομή $A \rightarrow B$ πραγματοποιούταν 10 φορές ενώ η $B \rightarrow A$ 4, θα αφαιρούσαμε την ακμή $B \rightarrow A$, αφαιρώντας ουσιαστικά όλα τα δεδομένα που απαρτίζονταν από αυτήν τη διαδρομή.

3.2.3 Δημιουργία Μπεϋζιανού δικτύου

Ένα μπεϋζιανό δίκτυο (B) εξαγόμενο από ένα αρχείο καταγραφής συμβάντων (E), θα μπορούσαμε να το ορίσουμε συνοπτικά ως $B_E = (\Delta, \Pi, \sigma)$, όπου:

- Δ , είναι το άκυκλο κατευθυνόμενο γράφημα
- Π , είναι ένα σύνολο πινάκων δεσμευμένων πιθανοτήτων, έναν για κάθε κόμβο του γραφήματος
- σ , μια συνάρτηση που αντιστοιχίζει τους πίνακες στους αντίστοιχους κόμβους

[11]

Από το αρχείο καταγραφής συμβάντων, εξάγουμε τους πίνακες δεσμευμένων πιθανοτήτων κάθε δραστηριότητας. Επίσης, από τα προηγούμενα βήματα έχουμε ένα άκυκλο κατευθυνόμενο γράφημα που ουσιαστικά πρόκειται για τον σκελετό του μπεϋζιανού μας δικτύου.

Στο παρόν βήμα δεν κάνουμε τίποτα άλλο από το να συνδυάσουμε τα δεδομένα που έχουμε από τους πίνακες με το γράφημα, δημιουργώντας έτσι το μπεϋζιανό δίκτυο.

3.2.4 Υποθέσεις και πρόβλεψη δραστηριότητας

Πρόκειται ουσιαστικά για το στάδιο της ανάλυσης, στο δεύτερο μισό της προτεινόμενης προσέγγισης. Σε αυτό το κεφάλαιο θα εφαρμόσουμε τον μαθηματικό τύπο Μπέϋζ (Bayes), θα «παίξουμε» με τις πιθανότητες και θα προβλέψουμε επόμενη δραστηριότητα μιας επιχειρησιακής διαδικασίας.

4. *Μελέτη Περίπτωσης: Αίτημα αποζημίωσης πτήσης σε αεροπορική εταιρία.*

Το αρχείο καταγραφής συμβάντων που χρησιμοποιήθηκε, πρόκειται για ένα από τα πολλά αρχεία καταγραφής συμβάντων που περιέχονται στο βιβλίο Process Mining - Data Science in Action του Wil van der Aalst [9]

(http://www.processmining.org/event_logs_and_models_used_in_book). Περιγράφεται ο χειρισμός ενός αιτήματος αποζημίωσης σε μια αεροπορική εταιρία. Οι πελάτες μπορούν να ζητήσουν αποζημίωση (είτε για πτήσεις που ακυρώθηκαν λόγω καιρού είτε για άλλες που καθυστέρησαν) καταχωρώντας το αίτημα και μέσα από ένα σύνολο διεργασιών καταλήγουμε στην απόφαση, ότι δικαιούνται αποζημίωση και πληρώνονται ή ότι δεν δικαιούνται κάποια αποζημίωση και ακυρώνεται το αίτημα.

Η υλοποίηση της περίπτωσης με βάση την προσέγγιση που αναφέραμε στο κεφάλαιο 3, γίνεται με τη χρήση της γλώσσας προγραμματισμού Python και οι τρεις βασικές βιβλιοθήκες που χρησιμοποιήθηκαν είναι οι pm4py (<https://pm4py.fit.fraunhofer.de>), pgmpy (<https://pgmpy.org>) και scikit-learn (<https://scikit-learn.org/stable/index.html>).

Η βιβλιοθήκη pm4py (process mining for python) είναι μια νέα βιβλιοθήκη εξόρυξης διαδικασιών. Υποστηρίζει τους αλγορίθμους εξόρυξης διαδικασιών καθώς και ενσωματώνει και άλλες γνωστές βιβλιοθήκες του επιστημονικού τομέα, όπως pandas, numpy, scipy και scikit-learn. Επιπλέον, όλες οι τεχνικές εξόρυξης διαδικασιών (ανακάλυψη διαδικασιών, έλεγχος συμμόρφωσης, αξιολόγηση μοντέλων, γραφήματα, διαχείριση περιπτώσεων κ.α.) υλοποιούνται από τη βιβλιοθήκη. [31]

Η βιβλιοθήκη pgmpy επιτρέπει τη διαχείριση πιθανοτικών γραφικών μοντέλων. Ο χρήστης μπορεί και δημιουργεί γραφικά μοντέλα και μέσω συμπερασμάτων να παίρνει απαντήσεις σε πιθανότητες ενδιαφέροντος. Υποστηρίζει αρκετά μοντέλα, όπως μπεϋζιανά δίκτυα, αλυσίδες Μαρκόβ, δυναμικά μπεϋζιανά δίκτυα και πολλά άλλα.

Τέλος, η βιβλιοθήκη scikit-learn περιέχει εργαλεία που αφορούν τον τομέα της μηχανικής μάθησης. Στην περίπτωση που μελετήθηκε, βοήθησε αρκετά στο κομμάτι της εκπαίδευσης αλλά και αξιολόγησης του πιθανοτικού μοντέλου.

4.1 Περί event log

Το αρχείο καταγραφής συμβάντων περιέχει 1391 περιπτώσεις, που προέρχονται από 21 διαφορετικές και το σύνολο των συμβάντων αθροιστικά ανέρχεται στα 7539. Επτά είναι τα βασικά μας συμβάντα-δραστηριότητες τα οποία και είναι: a – register request , b – examine thoroughly, c – examine casually, d – check ticket, e – decide, f – reinitiate request, g – pay compensation και h – reject request. Η αρχική μας κατάσταση είναι το συμβάν a -register request ενώ σαν τελικές καταστάσεις έχουμε τα g - pay compensation και h – reject request.

Ο πίνακας 1 παρακάτω συνοψίζει τις βασικές πληροφορίες του αρχείου καταγραφής συμβάντων ενώ ο πίνακας 2 μας δείχνει όλες τις πιθανές περιπτώσεις που μπορεί να συναντήσουμε. Τέλος, στον πίνακα 3 παρατηρούμε ένα μέρος του αρχείου καταγραφής συμβάντων σε μορφή xes.

Πίνακας 1. Βασικές πληροφορίες αρχείου καταγραφής συμβάντων.

Συμβάν	Αριθμός εμφάνισης
a (register request)	1391
b (examine thoroughly)	566
c (examine casually)	971
d (check ticket)	1537
e (decide)	1537
f (reinitiate request)	146
g (pay compensation)	461
h (reject request)	930

Πίνακας 2. Σύνολο των περιπτώσεων (a = register request, b = examine thoroughly, c = examine casually, d = check ticket, e = decide, f = reinitiate request, g = pay compensation, h = reject request).

Αριθμός εμφάνισης περίπτωσης	Αναγνωστικό περίπτωσης	Ακολουθία συμβάντων
455	Case 1	a c d e h
191	Case 2	a b d e g
177	Case 3	a d c e h
144	Case 4	a b d e h
111	Case 5	a c d e g
82	Case 6	a d c e g
56	Case 7	a d b e h
47	Case 8	a c d e f d b e h
38	Case 9	a d b e g
33	Case 10	a c d e f b d e h
14	Case 11	a c d e f b d e g
11	Case 12	a c d e f d b e g
9	Case 13	a d c e f c d e h
8	Case 14	a d c e f d b e h
5	Case 15	a d c e f b d e g
3	Case 16	a c d e f b d e f d b e g
2	Case 17	a d c e f d b e g
2	Case 18	a d c e f b d e f b d e g
1	Case 19	a d c e f d b e f b d e h
1	Case 20	a d b e f b d e f d b e g
1	Case 21	a d c e f d b e f c d e f d b e g

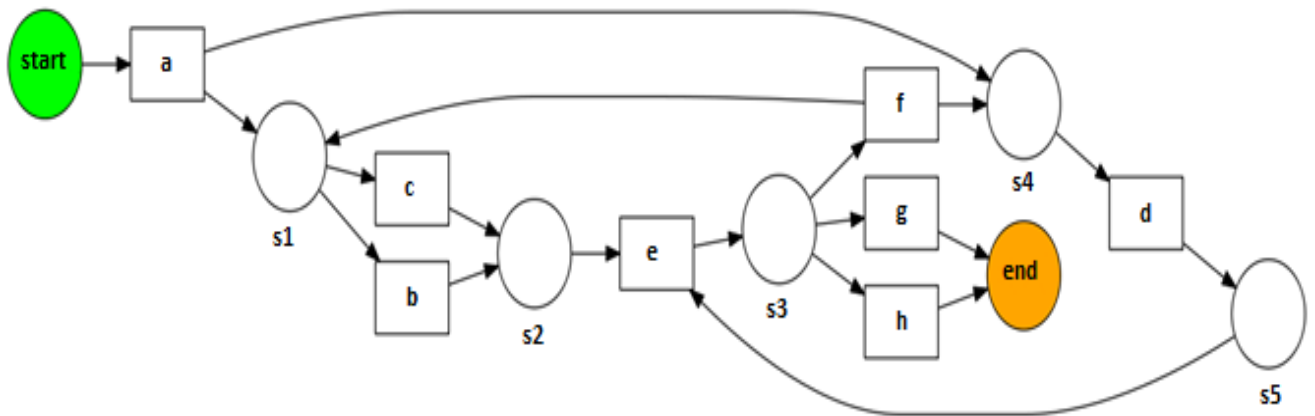
Πίνακας 3. Μέρος του αρχείου καταγραφής σε μορφή xes.

Αναγνωστικό περίπτωσης	Όνομα δραστηριότητας	Μετάβαση	Χρονική σήμανση
Case1.290	a	complete	2010-11-09 13:03:41.887000+01:00
Case1.290	c	complete	2010-11-09 13:04:41.887000+01:00
Case1.290	d	complete	2010-11-09 13:05:41.887000+01:00
Case1.290	e	complete	2010-11-09 13:06:41.887000+01:00
Case1.290	h	complete	2010-11-09 13:07:41.887000+01:00
Case1.289	a	complete	2010-11-09 13:03:41.887000+01:00
Case1.289	c	complete	2010-11-09 13:04:41.887000+01:00

4.2 Ανακάλυψη επιχειρησιακού μοντέλου

Θα ασχοληθούμε με τρεις αλγόριθμους ανακάλυψης διαδικασιών, τον Alpha Miner, Inductive Miner και Heuristic Miner. Επίσης, δεν ξεχνάμε μια νέα κατηγορία μοντέλου διαδικασιών, τα Directly-Follows Graphs. Θα γίνει μια παρουσίαση των παραπάνω μοντέλων και αργότερα στην πορεία θα τα συγκρίνουμε με στόχο να βρούμε το μοντέλο που αντιπροσωπεύει την περίπτωσή μας καλύτερα.

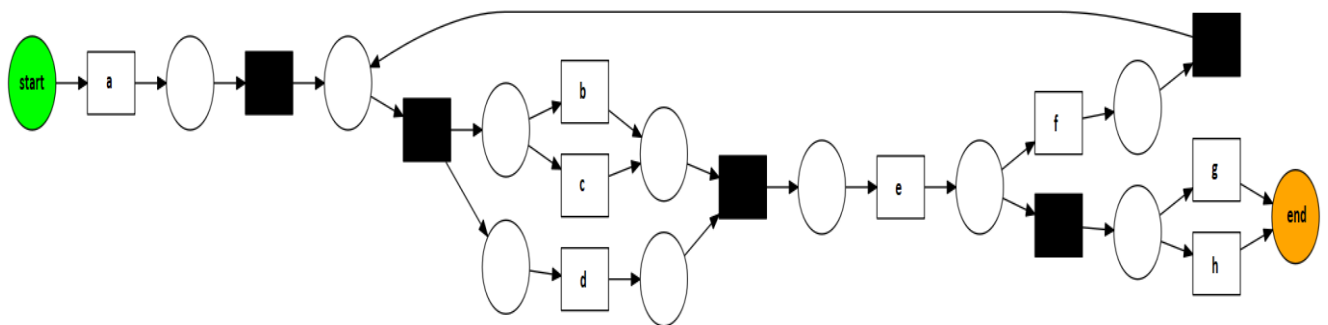
Ξεκινώντας με τον αλγόριθμο Alpha miner, το παραγόμενο μοντέλο διεργασιών σε μορφή γραφήματος Petri net φαίνεται στην εικόνα 25.



Εικόνα 25. Μοντέλο διεργασιών παραγόμενο από τον Alpha Miner (a = register request, b = examine thoroughly, c = examine casually, d = check ticket, e = decide, f = reinstantiate request, g = pay compensation, h = reject request).

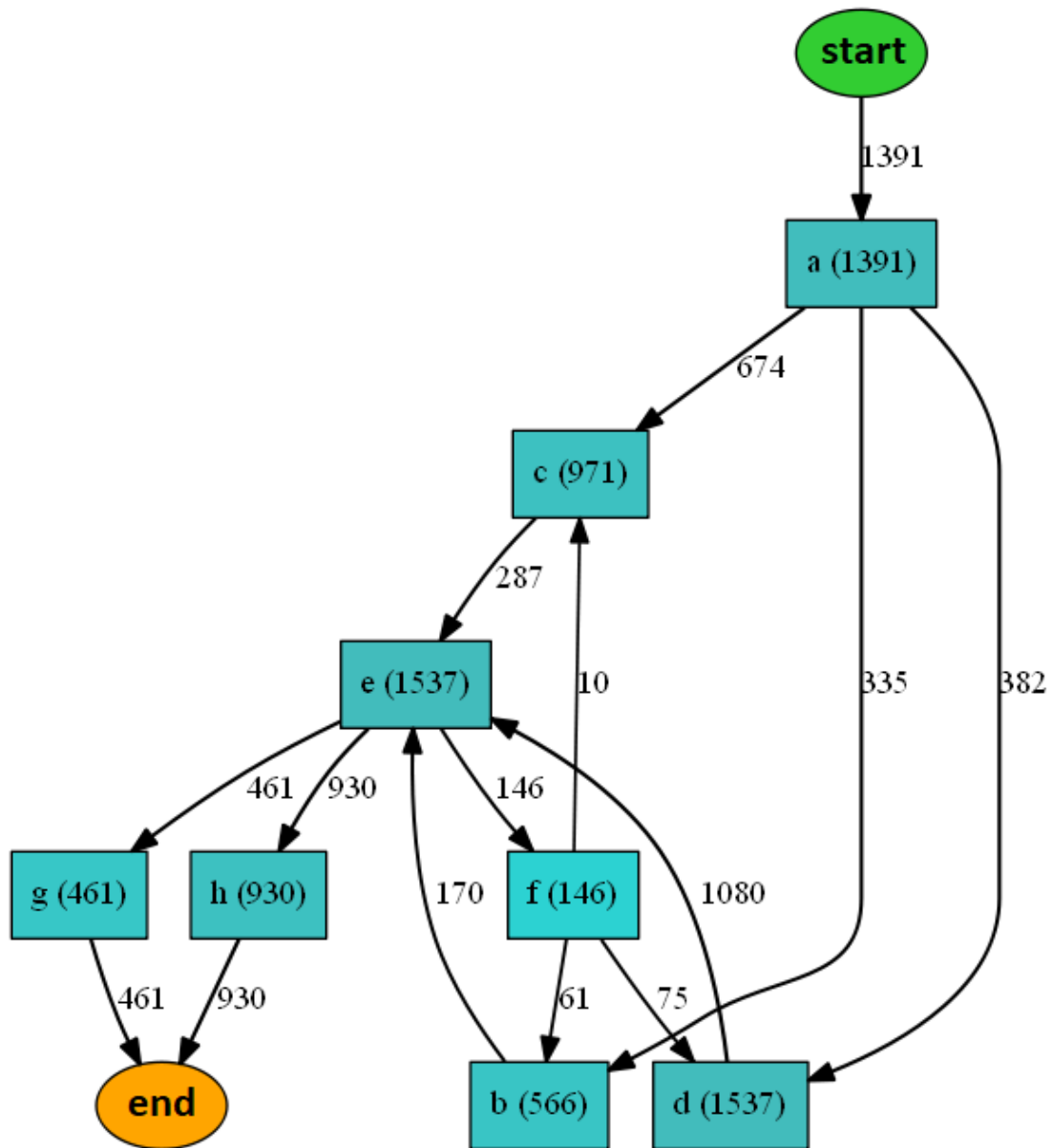
Οι δραστηριότητες-συμβάντα του αρχείου καταγραφής αντιπροσωπεύονται με τετράγωνα, συχνά αποκαλούνται και μεταβάσεις. Αυτές οι μεταβάσεις συνδέονται μεταξύ τους μέσω κύκλων (s1, s2, ..., s5), δηλαδή κάποιες τοποθεσίες που μοντελοποιούν πιθανές καταστάσεις μιας διαδικασίας. [9, 13]

Συνεχίζουμε με τον αλγόριθμο Inductive Miner στην εικόνα 26.



Εικόνα 26. Μοντέλο διεργασιών παραγόμενο από τον Inductive Miner (a = register request, b = examine thoroughly, c = examine casually, d = check ticket, e = decide, f = reinstantiate request, g = pay compensation, h = reject request).

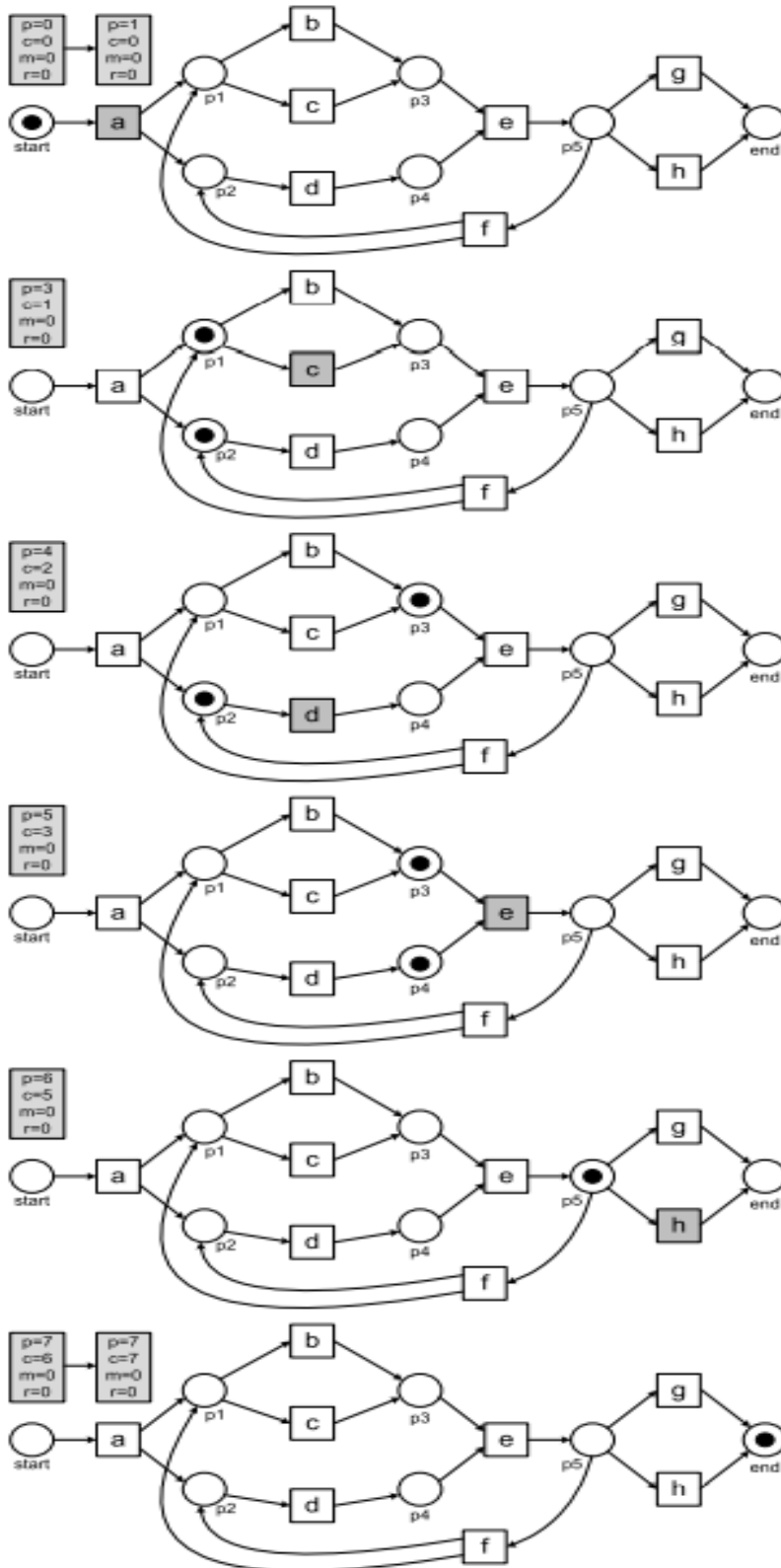
Τα μαύρα χρωματισμένα τετράγωνα, οι μεταβάσεις, ονομάζονται σιωπηλές ή άορατες μεταβάσεις και δεν αντιπροσωπεύουν καμία δραστηριότητα του αρχείου καταγραφής συμβάντων. Είναι αναγκαίες όμως για τη μοντελοποίηση κάποιας σχέσης μεταξύ των δραστηριοτήτων (σχέσεις παραλληλίας ή αποκλειστικότητας). Αντίστοιχα με AND/OR



Εικόνα 28. Μοντέλο διεργασιών Directly-Follows Graph (a = register request, b = examine thoroughly, c = examine casually, d = check ticket, e = decide, f = reiniciate request, g = pay compensation, h = reject request).

4.3 Έλεγχος συμμόρφωσης στην πράξη

Όπως αναφέραμε και στο κεφάλαιο 3, θα εξετάσουμε δύο τεχνικές συμμόρφωσης. Ξεκινώντας με την τεχνική Token-based replay, αναπαράγουμε την κάθε περίπτωση που συναντάμε στο αρχείο καταγραφής συμβάντων πάνω σε κάθε επιχειρησιακό μοντέλο. Μια τέτοια αναπαράσταση φαίνεται με βήματα στην εικόνα 29:



Εικόνα 29. Replaying the case a-c-d-e-h. (p = produced tokens, c = consumed tokens, m = missing tokens, r = remaining tokens).

Ο γενικός τύπος με τον οποίο θα ελέγξουμε την προσαρμογή (fitness) είναι ο εξής:

$$\text{Fitness}(\mathbf{n}, \mathbf{t}) = \frac{1}{2} * [(1 - \frac{r}{p}) + (1 - \frac{m}{c})],$$

όπου δέχεται ως ορίσματα ένα επιχειρησιακό μοντέλο σε μορφή petri net (n) και μια περίπτωση (t) από το αρχείο καταγραφής συμβάντων. Για κάθε περίπτωση συναντάμε 4 μεταβλητές, p, r, m, c, produced tokens, remaining tokens, missing tokens και consumed tokens αντίστοιχα. [9,24]

Ξεκινάμε με τη μεταβλητή $p = 1$ καθώς η αρχή (start) παράγει ένα token και όλες οι άλλες μεταβλητές ισούνται με μηδέν. Κάθε φορά που φτάνουμε και εκτελούμε μια δραστηριότητα καταναλώνουμε ένα token, αυξάνοντας τον αριθμό c. Προχωρώντας από τη δραστηριότητα a παράγουμε 2 tokens (αφού έχουμε δυο ακμές), αυξάνουμε τον αριθμό p και ούτω καθεξής. Αν μια δραστηριότητα δεν έχει token για να εκτελεστεί, προσθέτουμε εμείς ένα token και αυξάνουμε τον αριθμό m κατά 1. Τέλος, αν φτάσουμε στο τέλος (end) και έχουμε tokens που περισσεύουν, αυξάνουμε τον αριθμό της μεταβλητής r ανάλογα. Στην περίπτωσή μας, στο μοντέλο του alpha miner, οι τέσσερις μεταβλητές (p, r, m, c) για την πρώτη περίπτωση του αρχείου καταγραφής συμβάντων, φαίνονται ως εξής: $m = 0, c = 7, r = 0, p = 7$. Εφαρμόζοντας τον τύπο της προσαρμογής που είδαμε παραπάνω, βγάζουμε $\text{Fitness} = 1$, δηλαδή η πρώτη περίπτωση αναπαράγεται τέλεια στο μοντέλο του alpha miner. Αν κάνουμε την ίδια διαδικασία για όλες τις περιπτώσεις, θα πάρουμε τον συνολικό αριθμό της προσαρμογής (fitness). Στην περίπτωση του alpha miner, η συνολική προσαρμογή ήταν 1 ([0-1] όσο μεγαλύτερη η τιμή τόσο το καλύτερο).

Στον παρακάτω πίνακα, βλέπουμε αναλυτικά για κάθε μοντέλο το βαθμό με τον οποίο μπορούν να αναπαράγουν τις περιπτώσεις του αρχείου καταγραφής συμβάντων (το directly follows graph δεν αποτελεί αποδεκτή είσοδο καθώς δεν είναι petri net και για αυτό απουσιάζει από τον πίνακα).

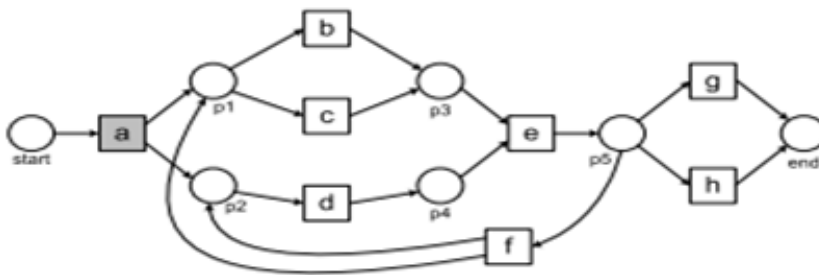
Πίνακας 4. Αποτελέσματα token-based συμμόρφωσης.

	fitness	perfect fit traces
alpha miner	1	100%
inductive miner	1	100%
heuristic miner	0.97	60%

Περνάμε τώρα στην τεχνική των Alignments. Μια τεχνική πιο εύκολα κατανοήσιμη από την token-based replay και θα δούμε παρακάτω το γιατί.

Πρόκειται για έναν πίνακα με 2 γραμμές, μια που αντιπροσωπεύει το αρχείο καταγραφής συμβάντων και μια το επιχειρησιακό μοντέλο.

Έστω ότι έχουμε την περίπτωση a-c-d-e-h, όπως είχαμε και παραπάνω και έστω το μοντέλο που θα χρησιμοποιήσουμε φαίνεται στην παρακάτω εικόνα.



Εικόνα 30. Alignments μοντέλο παραδείγματος.

Ξεκινάμε με τη δραστηριότητα a και ανάλογα με τη σειρά που παρατηρούνται οι δραστηριότητες στην περίπτωση, αν μπορούν να αναπαρασταθούν από το μοντέλο συμπληρώνονται στον πίνακα με τη σειρά. Συνεχίζουμε με την c, την οποία και παρατηρούμε στο μοντέλο, οπότε συμπληρώνεται ο πίνακας ανάλογα. Επόμενη δραστηριότητα στην περίπτωση είναι η d , η οποία μπορεί και εκτελείται από το μοντέλο, οπότε συμπληρώνουμε τον πίνακα και ούτω καθεξής μέχρι να φτάσουμε στο τέλος. Οπότε συμπληρώνουμε τη γραμμή του μοντέλου στον πίνακα 5 με τη δραστηριότητα που εκτελείται τη φορά.

Πίνακας 5. Alignments παράδειγμα 1.

περίπτωση	a	c	d	e	h
μοντέλο	a	c	d	e	h

Μια άλλη περίπτωση που αξίζει να εξετάσουμε είναι όταν μια περίπτωση από το αρχείο καταγραφής συμβάντων, δεν μπορεί να αναπαραχθεί πλήρως από το επιχειρησιακό μοντέλο.

Έστω ότι η περίπτωση μας αποτελείται από τις εξής δραστηριότητες: a-c-e-h.

Ο πίνακας 6 περιγράφει τη τελική μορφή που θα παίρναμε. Ξεκινάμε με τα a και c, τα οποία μπορούν να αναπαραχθούν από το μοντέλο. Στη συνέχεια όμως στη περίπτωση μας ακολουθεί η δραστηριότητα e, ενώ στο μοντέλο εκτελείται η d. Συμπληρώνουμε τη γραμμή του μοντέλου με τη δραστηριότητα d και στη γραμμή της περίπτωσης προσθέτουμε '>>' και σημαίνει ότι δεν έχουμε προχωρήσει ακόμα στη περίπτωση. Έπειτα προχωράμε στην επόμενη δραστηριότητα του αρχείου καταγραφής μέχρι το τέλος. Έτσι, ουσιαστικά προχωράμε μόνο στο μοντέλο ενώ η σειρά των δραστηριοτήτων στην περίπτωση μένει στάσιμη. Μπορούμε να παρατηρήσουμε το ίδιο αλλά με την ανάποδη λογική, δηλαδή να μείνει στάσιμο το μοντέλο και να προχωρήσει η περίπτωση.

Πίνακας 6. Alignments παράδειγμα 2.

περίπτωση	a	c	>>	e	h
μοντέλο	a	c	d	e	h

Με τα παραπάνω, μετράμε το κόστος στην τεχνική των alignments. Αν ένα κελί δεν είναι όμοιο και στην περίπτωση και στο μοντέλο, αυξάνουμε το κόστος κατά ένα. Μικρότερο κόστος ισοδυναμεί με καλύτερη αναπαράσταση του αρχείου καταγραφής συμβάντων από το επιχειρησιακό μοντέλο.

Για τον υπολογισμό της προσαρμογής (fitness), έχουμε τον εξής τύπο:

$$Fitness = 1 - \left[\frac{(alignments\ cost)}{model\ cost + trace\ cost} \right],$$

όπου alignments cost είναι το κόστος που παρατηρούμε όταν αναπαράγουμε μια περίπτωση και έχουμε διαφορετικές τιμές στα κελιά, ενώ model και trace cost ουσιαστικά αποτελούν το πλήθος των ενεργειών που θα εκτελέσει το μοντέλο και το πλήθος των δραστηριοτήτων στην περίπτωση αντίστοιχα. Το άθροισμα του αριθμού fitness για κάθε μια περίπτωση διαιρεμένο με το πλήθος των συνολικών περιπτώσεων που παρατηρούνται στο αρχείο καταγραφής συμβάντων αποτελεί τον αριθμό της συνολικής προσαρμογής για το μοντέλο.

Παραθέτεται παρακάτω ο πίνακας με τα αποτελέσματα για τα μοντέλα που έχουμε στην περίπτωση μας:

Πίνακας 7. Αποτελέσματα alignments συμμόρφωσης.

	fitness	perfect fit traces
alpha miner	1	100%
inductive miner	1	100%
heuristic miner	0.92	60%

[9,24]

4.4 Αξιολόγηση των επιχειρησιακών μοντέλων

Τα τέσσερα κριτήρια που θα συγκριθούν είναι: η προσαρμογή (fitness), η απλότητα (simplicity), η γενίκευση (generalization) και η ακρίβεια (precision).

Για την προσαρμογή και την ακρίβεια, υπάρχουν δύο τρόποι εξέτασης, μέσω token-based replay και alignments. Σε γενικές γραμμές η βασική ιδέα των δύο αυτών τρόπων είναι ίδια, η μόνη διαφορά τους είναι ότι η token-based replay είναι γρηγορότερη αλλά βασίζεται σε ευρετικές μεθόδους αναζήτησης (επομένως και το αποτέλεσμα μπορεί να μην είναι απόλυτα ακριβές) ενώ με τα alignments παίρνουμε έγκυρα αποτελέσματα αλλά μπορεί να καθυστερήσουν εάν το μοντέλο είναι μεγάλο. [14, 15]

Στους παρακάτω πίνακες βλέπουμε αναλυτικά τα αποτελέσματα. Οι τιμές κυμαίνονται μεταξύ 0 και 1 (όσο μεγαλύτερη τιμή τόσο καλύτερα).

Πίνακας 8. Αξιολόγηση μοντέλου Alpha Miner.

Alpha miner		
Fitness	<i>Using Token-based replay</i>	<i>Using Alignments</i>
	1	1
Precision	<i>Using Token-based replay</i>	<i>Using Alignments</i>
	0.95	0.95
Simplicity	0.65	
Generalization	0.96	
Overall score	0.89	

Πίνακας 9. Αξιολόγηση μοντέλου Inductive Miner.

Inductive miner		
Fitness	<i>Using Token-based replay</i>	<i>Using Alignments</i>
	1	1
Precision	<i>Using Token-based replay</i>	<i>Using Alignments</i>
	0.95	0.95
Simplicity	0.80	
Generalization	0.96	
Overall score	0.93	

Πίνακας 10. Αξιολόγηση μοντέλου Heuristic Miner.

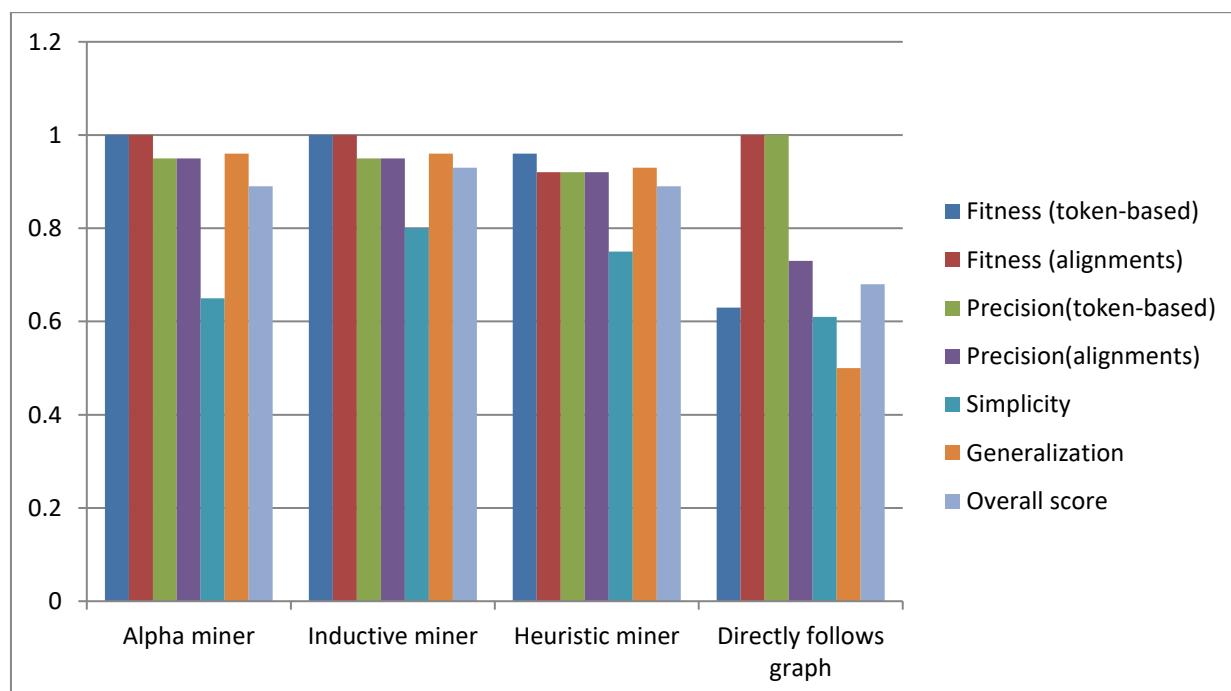
Heuristic miner		
Fitness	<i>Using Token-based replay</i>	<i>Using Alignments</i>
	0.96	0.92
Precision	<i>Using Token-based replay</i>	<i>Using Alignments</i>
	0.92	0.92
Simplicity	0.75	
Generalization	0.93	
Overall score	0.89	

Πίνακας 11. Αξιολόγηση μοντέλου Directly-Follows Graph.

Directly Follows Graph		
Fitness	<i>Using Token-based replay</i>	<i>Using Alignments</i>
	0.63	1
Precision	<i>Using Token-based replay</i>	<i>Using Alignments</i>
	1	0.73
Simplicity	0.61	
Generalization	0.5	
Overall score	0.68	

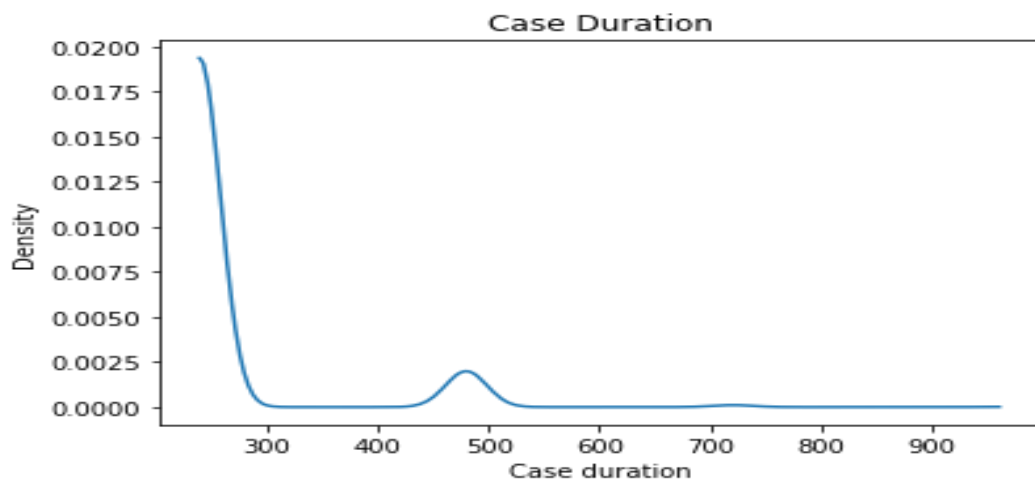
Μεταξύ των Alpha miner, Inductive miner, Heuristic miner και Directly follows graph, το μοντέλο διεργασιών που εξήγε ο Inductive miner πέραν ότι είχε το καλύτερο μέσο όρο συνολικά αποτέλεσμα, είχε και το καλύτερο σκορ σε κάθε ένα από τα τέσσερα κριτήρια που ορίσαμε για σύγκριση. Επομένως, το μοντέλο του ανταποκρίνεται καλύτερα στις απαιτήσεις της περίπτωσης μας.

Πίνακας 12. Σύγκριση των μοντέλων.

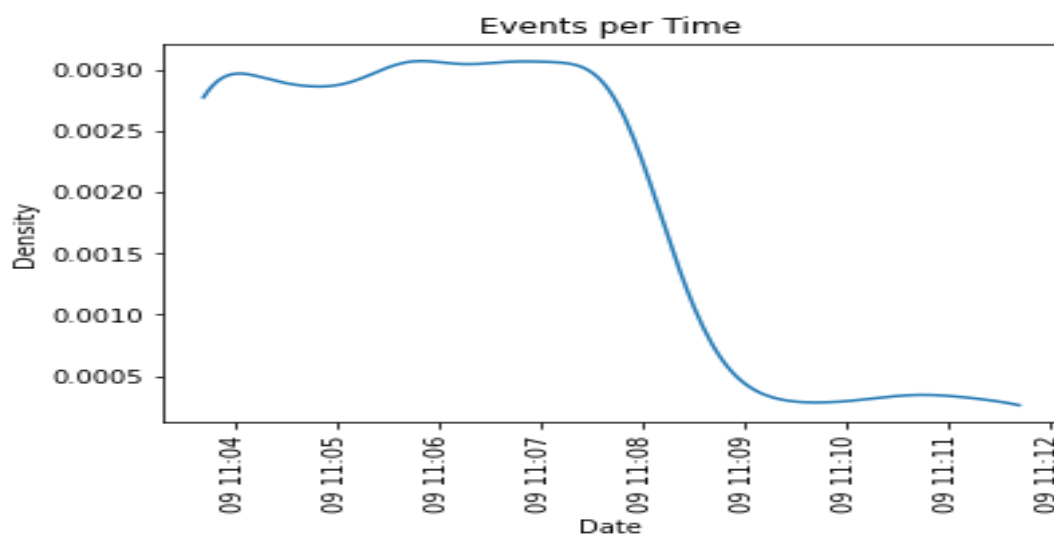


4.5 Λοιπές πληροφορίες για το αρχείο καταγραφής συμβάντων (statistics, decision mining)

Μέσα από κάποιο αρχείο καταγραφής συμβάντων μπορούμε να εξάγουμε και κάποια στατιστικά στοιχεία του, χωρίς να μας είναι απαραίτητα για κάποια βασική λειτουργία της προτεινόμενης προσέγγισης, αλλά για να έχουμε μια πιο γενική εικόνα. Έτσι, βλέπουμε παρακάτω μερικά στοιχεία που αφορούν ένα από τα χαρακτηριστικά του αρχείου καταγραφής, τη χρονική σήμανση (timestamp).



Εικόνα 31. Κατανομή χρονικής διάρκειας μιας περίπτωσης.



Εικόνα 32. Κατανομή των δραστηριοτήτων ανά την πάροδο του χρόνου.

Πίνακας 13. Χρονικές λεπτομέρειες του αρχείου καταγραφής συμβάντων.

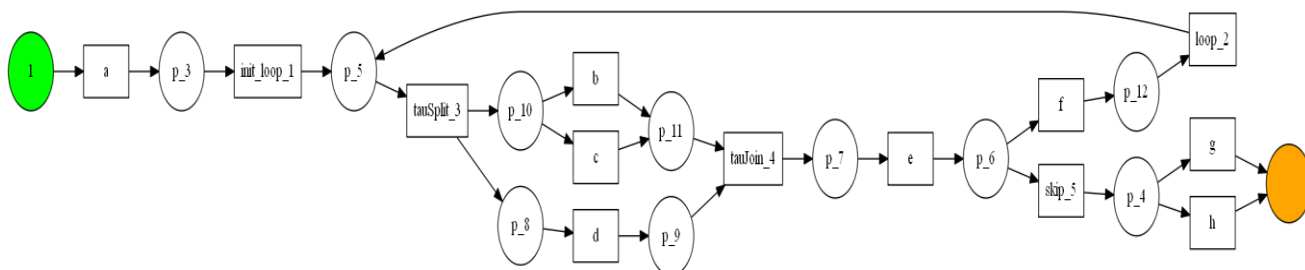
Χρονική διάρκεια (λεπτά και δευτερόλεπτα)	
Average duration of a case	4:41
Median value of a case	4

Όπως είδαμε και παραπάνω στο μοντέλο του Inductive miner (εικόνα 26), παρατηρήσαμε κάποια σημεία τα οποία είχαν δύο ακμές προς μια μετάβαση. Αυτά τα σημεία ονομάζομαι σημεία απόφασης. Συγκεκριμένα, τα δύο αυτά σημεία είναι :

- μετά από τη μετάβαση a (register request) και προς τα b, c (examine thoroughly και examine casually αντίστοιχα).
- μετά από τη μετάβαση e (decide) προς τα f (reinitiate request) και g (pay compensation) και h (reject request).

Σε κάποιο BPMN αυτά τα σημεία αντιστοιχούν σε κάποια λογική πύλη OR, όπως και αναφέραμε παραπάνω και πρόκειται για ένα σημείο το οποίο θα επιλέξει μια διαδρομή μεταξύ των δύο που έχει για να συνεχίσει. Η εξόρυξη αποφάσεων που θα εφαρμόσουμε σε κάποιο σημείο για να δούμε ποια διαδρομή θα επιλέξει, πρόκειται για μια υποκατηγορία των δέντρων απόφασης (χρησιμοποιούνται συχνά για αναζήτηση ανωμαλιών σε ένα επιχειρησιακό μοντέλο) και δεν θα επιμείνουμε, παρά μόνο αναφέρεται σαν μια επιπρόσθετη τεχνική της εξόρυξης διεργασιών. [9]

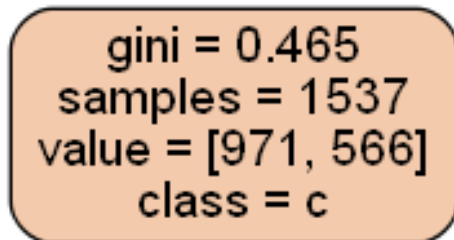
Δημιουργήθηκε ένα μοντέλο βασισμένο στον Inductive miner algorithm και μπορούμε να δούμε με ευκολία τα σημεία απόφασης (decision points) στην παρακάτω εικόνα.



Εικόνα 33. Decision points on Inductive miner process model.

Θα ασχοληθούμε με το σημείο 'p10', ένα σημείο απόφασης μεταξύ των δραστηριοτήτων b (examine throughly) και c (examine casually).

Εκτελώντας τον αλγόριθμο εξόρυξης αποφάσεων, το αποτέλεσμα που παίρνουμε φαίνεται στην εικόνα 34.



Εικόνα 34. Decision mining results.

Στο σημείο 'p10' βλέπουμε ότι υπάρχουν συνολικά 1537 δείγματα, 971 που αντιπροσωπεύουν τη δραστηριότητα c (examine casually) και 566 για την b (examine throughly). Η απόφαση να ακολουθήσει τη δραστηριότητα c πάρθηκε βάσει του δείκτη τζίνι (Gini index), όπως και βλέπουμε από την εικόνα με τα αποτελέσματα.

4.6 Προετοιμασία για τη δημιουργία Μπεϋζιανού μοντέλου

Από αυτό το σημείο και παρακάτω, περνάμε ουσιαστικά στα αποτελέσματα του δεύτερου μέρους της προτεινόμενης προσέγγισης.

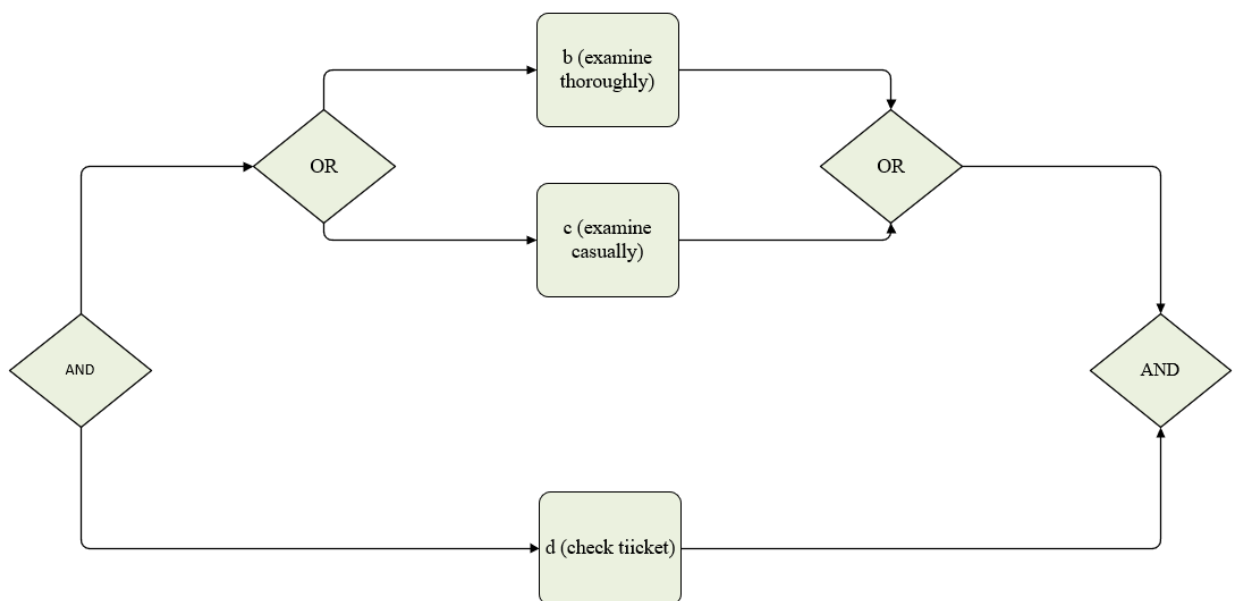
Πρωταρχικό βήμα είναι τα μετατρέψουμε κάποιο από τα επιχειρησιακά μοντέλα που εξάγαμε παραπάνω σε κάποιο νέο άκυκλο μοντέλο.

Ήδη από το πρώτο μισό μέρος της προσέγγισης έχουμε έτοιμα τα επιχειρησιακά μοντέλα. Για την περίπτωση που εξετάζουμε, προτιμήθηκε το μοντέλο του Directly follows graph μιας και απεικονίζει μόνο τις δραστηριότητες και τις σχέσεις των και είναι εύκολα αναγνώσιμο.

4.7 Αναζήτηση κύκλων

Αρχικά ελέγχουμε το μοντέλο μας με κάποιον αλγόριθμο εντοπισμού κύκλων. Αν το μοντέλο δεν περιέχει κύκλους μέσα στις δραστηριότητές του, μπορούμε να προσπεράσουμε το επόμενο βήμα (αφαίρεση κύκλων).

Χρησιμοποιώντας κάποιον αλγόριθμο εντοπισμού κύκλων παρατηρούμε ότι κανένα από τα μοντέλα δεν είναι άκυκλο καθώς η δραστηριότητα *f* (reinitiate request) δημιουργεί ένα κύκλο από μόνη της (πίνακας 14). Πέραν αυτού, όπως βλέπουμε στην εικόνα 33, αν παρατηρήσουμε το κομμάτι μεταξύ των μεταβάσεων 'tauSplit_3' και 'tauJoin_4' θα μπορούσε κανείς να βγάλει το συμπέρασμα ότι γίνεται πάντα η δραστηριότητα *d*, σε συνδυασμό με την *b* και *c*, μια από τις δύο τη φορά. Το αντίστοιχο επιχειρησιακό μοντέλο φαίνεται στην παρακάτω εικόνα σε μορφή BPMN.



Εικόνα 35. BPMN ενός μέρους του μοντέλου εξαγόμενο από τον inductive miner.

Η αναπαράσταση των δύο λογικών πυλών (and/και, or/ή) σε γράφημα δημιουργούν ακόμα ένα κύκλο, με αποτέλεσμα να δημιουργηθούν ακόμα δύο σειρές κύκλων με επίκεντρο το *d*, τα *d-c-c-d* και *d-b-b-d*. Το σύνολο των κύκλων φαίνεται παρακάτω στον πίνακα 14.

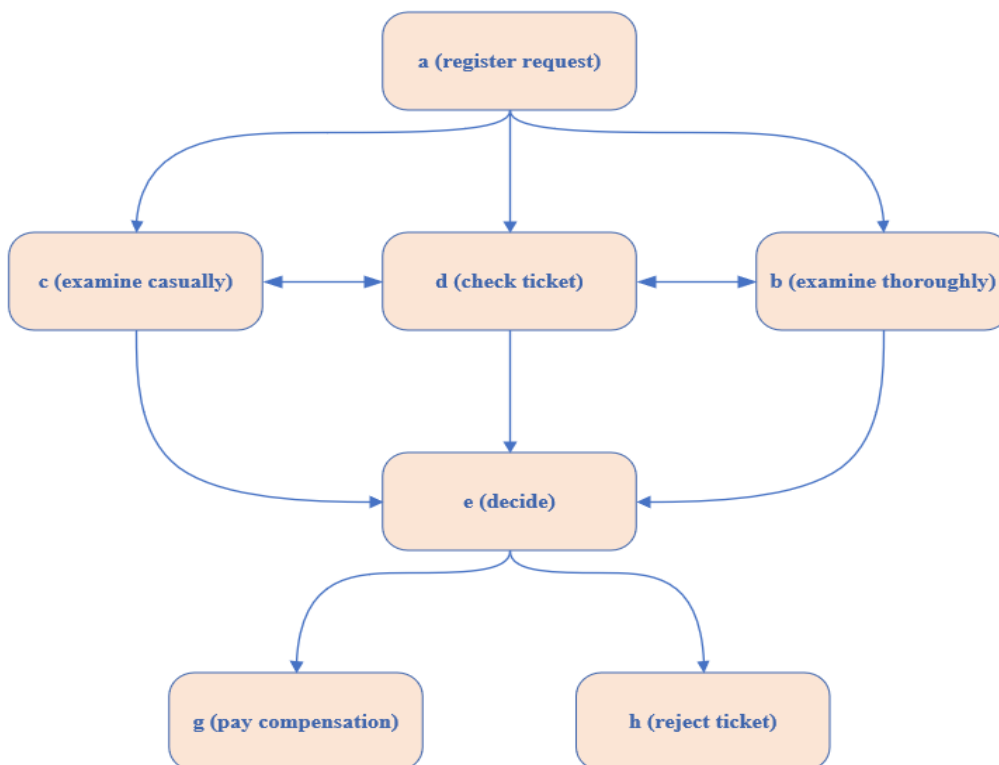
Πίνακας 14. Ακολουθία δραστηριοτήτων που ανήκουν σε κύκλο.

Ακολουθία δραστηριοτήτων που ανήκουν σε κύκλο	
Alpha miner	[f, d, e], [f, c, e], [f, b, e]
Inductive miner	[b, e, f, loop_2], [c, e, f, loop_2], [d, e, f, loop_2]
Heuristic miner	[e, f, d], [e, f, b], [e, f, c]
Directly follows graph	[f, b, e], [f, b, d, e], [f, b, d, c, e], [f, d, e], [f, d, b, e], [f, d, c, e], [f, c, d, e], [f, c, d, b, e], [f, c, e], [d, b], [c, d]

Έτσι λοιπόν, περνάμε στο επόμενο βήμα, την αφαίρεση κύκλων.

4.8 Αφαίρεση κύκλων

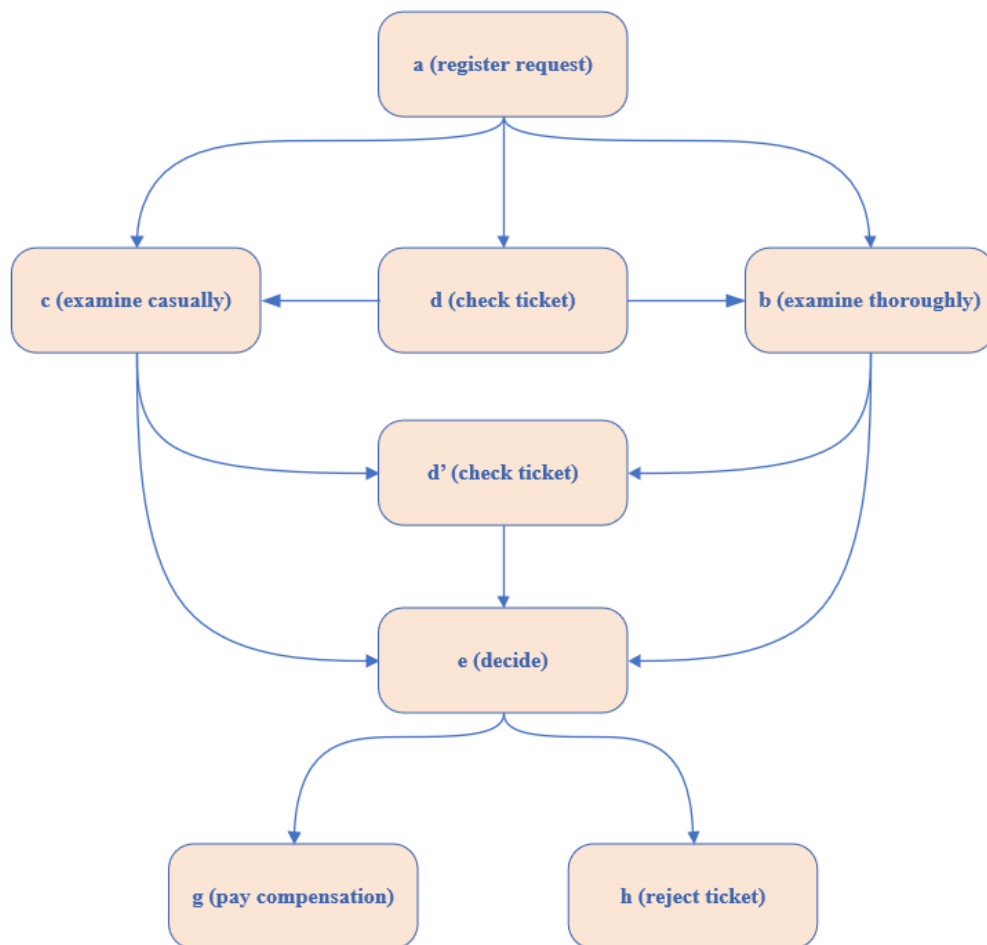
Πρώτο μας μέλημα είναι να αφαιρέσουμε από το αρχείο καταγραφής συμβάντων τη δραστηριότητα f. Το νέο μοντέλο διεργασιών, με γνώμονα το Directly follows graph, φαίνεται παρακάτω στην εικόνα 36.



Εικόνα 36. Εκ νέου Directly follows graph.

Όπως παρατηρούμε μένει να αφαιρέσουμε το σετ κύκλων d-c και d-b. Σε αυτό το σημείο θα εφαρμόσουμε την τεχνική αφαίρεση κύκλων που αναφέραμε στην προσέγγιση, δηλαδή αντιπροσώπευση κύκλου με πολλές εμφανίσεις του ίδιου κόμβου.

Έπειτα από την αφαίρεση των κύκλων, το τελικό άκυκλο κατευθυνόμενο γράφημα που θα χρησιμοποιήσουμε και παρακάτω για τη δημιουργία του μπεϋζιανού φαίνεται στην εικόνα 37.



Εικόνα 37. Directed Acyclic Graph (DAG).

4.9 Κατασκευή Μπεϋζιανού μοντέλου

Επόμενο και τελευταίο βήμα πριν την ολοκλήρωση του μπεϋζιανού δικτύου είναι ο υπολογισμός των πινάκων δεσμευμένων πιθανοτήτων κάθε δραστηριότητας.

Πριν όμως προχωρήσουμε στην εξαγωγή των πινάκων αυτών, ας δούμε πως μοιάζει πλέον το αρχείο καταγραφής συμβάντων, έπειτα από την αφαίρεση της δραστηριότητας f (reinitiate request) και την προσθήκη ενός νέου κόμβου (d' – check ticket).

Το σύνολο των περιπτώσεων έπεσε στα 1254, 137 περιπτώσεις λιγότερες από προηγουμένως, ενώ ο αριθμός των συνολικών συμβάντων έπεσε στα 6270 (-1269).

Πίνακας 15. Βασικές πληροφορίες φιλτραρισμένου αρχείου καταγραφής συμβάντων.

Συμβάν	Αριθμός εμφάνισης
a (register request)	1254
b (examine thoroughly)	429
c (examine casually)	825
d (check ticket)	353
d' (check ticket)	901
e (decide)	1254
g (pay compensation)	422
h (reject request)	832

Πίνακας 16. Πίνακας 2. Σύνολο των περιπτώσεων φιλτραρισμένου αρχείου καταγραφής συμβάντων.

Αριθμός εμφάνισης περίπτωσης	Αναγνωστικό περίπτωσης	Ακολουθία συμβάντων
455	Case 1	a c d' e h
191	Case 2	a b d' e g
177	Case 3	a d c e h
144	Case 4	a b d' e h
111	Case 5	a c d' e g
82	Case 6	a d c e g
56	Case 7	a d b e h
38	Case 9	a d b e g

Κατά τη δημιουργία του μπεϋζιανού δικτύου, θα χρειαστεί να χωρίσουμε το αρχείο καταγραφής συμβάντων σε δύο μέρη με τυχαίο τρόπο, ένα για να εκπαιδεύσουμε το μοντέλο μας και ένα για να το εξετάσουμε. Ο λόγος εκπαίδευσης-εξέτασης θα μπορούσε να οριστεί αριθμητικά ως 80% - 20%.

Οι πίνακες δεσμευμένων πιθανοτήτων (CBT) που προκύπτουν από το μέρος του αρχείου καταγραφής που χρησιμοποιήθηκε για να εκπαιδευτεί το μοντέλο μας είναι οι εξής:

Πίνακας 17. CBT of a.

a(1)	1
------	---

Πίνακας 18. CBT of b.

a	a(1)	a(1)
d	d(0)	d(1)
b(0)	0.612	0.754
b(1)	0.387	0.245

Πίνακας 19. CBT of c.

a	a(1)	a(1)
d	d(0)	d(1)
c(0)	0.387	0.245
c(1)	0.612	0.754

Πίνακας 20. CBT of d.

a	a(1)
d(0)	0.715
d(1)	0.284

Πίνακας 21. CBT of d'.

b	b(0)	b(0)	b(1)	b(1)
c	c(0)	c(0)	c(1)	c(1)
d'(0)	0.5	0.328	0.201	0.5
d'(1)	0.5	0.671	0.798	0.5

Πίνακας 22. CBT of e.

b	b(0)	b(0)	b(0)	b(0)	b(1)	b(1)	b(1)	b(1)
c	c(0)	c(0)	c(0)	c(0)	c(1)	c(1)	c(1)	c(1)
d'	d'(0)	d'(0)	d'(0)	d'(0)	d'(1)	d'(1)	d'(1)	d'(1)
e(1)	1	1	1	1	1	1	1	1

Πίνακας 23. CBT of output activity (g/h).

e	e(1)
g(1)	0.339
h(1)	0.660

Όπου (1) η τιμή εμφάνισης της δραστηριότητας και (0) η τιμή απουσίας αντίστοιχα. Όπως είδαμε και στη θεωρία, οι πίνακες είναι ίσοι σε αριθμό με τις δραστηριότητες του αρχείου καταγραφής και ουσιαστικά περιγράφουν την τιμή εμφάνισης μιας δραστηριότητας, δεδομένου της εμφάνισης των 'πατρικών' τους δραστηριοτήτων. Σε αυτό το σημείο το μπεϋζιανό δίκτυο έχει ολοκληρωθεί, αφού έχουμε και άκυκλο κατευθυνόμενο γράφημα (εικόνα 38) και τους πίνακες δεσμευμένων πιθανοτήτων (CBT) (πίνακες 17 έως 23).

4.10 Ανάλυση και αποτελέσματα

Διαλέξαμε τα μπεϋζιανά δίκτυα ως ένα τρόπο διαχείρισης της αβεβαιότητας μεταξύ επιχειρησιακών διεργασιών. Και αυτό διότι μια από τις ιδιότητές τους μας επιτρέπει να υπολογίσουμε το ποσοστό εμφάνισης μιας επιχειρησιακής δραστηριότητας, γνωρίζοντας πως έχουν προηγηθεί, από μερικές έως και καμία, άλλες δραστηριότητες στην επιχειρησιακή διαδικασία.

Προφανώς, όσες λιγότερες δραστηριότητες γνωρίζουμε ότι έχουν συμβεί, τόσο μεγαλύτερη είναι και η αβεβαιότητα για τον υπολογισμό εμφάνισης μιας άλλης δραστηριότητας. Παρακάτω, παραθέτεται ένας πίνακας με αποτελέσματα από μερικές ενδεικτικές δοκιμές, στην προσπάθεια υπολογισμού μιας πιθανότητας εμφάνισης μιας δραστηριότητας ενώ έχει παρατηρηθεί μια άλλη (ουσιαστικά εφαρμογή του τύπου Bayes).

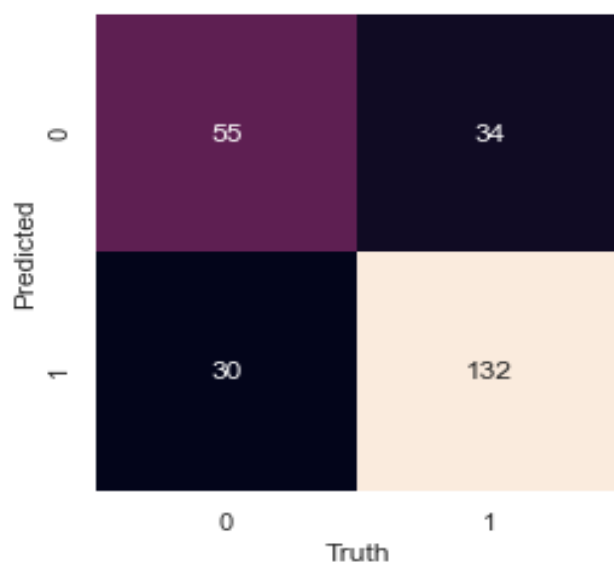
Πίνακας 24. Υπολογισμός πιθανοτήτων εμφάνισης μιας δραστηριότητας ενώ γνωρίζουμε ότι έχει εκτελεστεί μια άλλη δραστηριότητα.

$\Pr(b = \text{yes} \mid d' = \text{yes})$	0.3386
$\Pr(d' = \text{yes} \mid e = \text{yes})$	0.6145
$\Pr(c = \text{yes} \mid d = \text{yes})$	0.7367
$\Pr(c = \text{yes}, e = \text{yes} \mid d = \text{yes})$	0.7328
$\Pr(e = \text{yes} \mid d' = \text{yes})$	1.0000
$\Pr(d' = \text{yes} \mid b = \text{yes})$	0.5884

Τέλος, έγινε προσπάθεια για την πρόβλεψη της τελικής δραστηριότητας στην περίπτωση που μελετάμε (g-pay compensation ή h-reject request). Το μπεϋζιανό μοντέλο που δημιουργήθηκε, χρησιμοποίησε το μέρος του αρχείου καταγραφής συμβάντων που ορίσαμε ως μέρος εκπαίδευσης για να μάθει τις συνδυαστικές πιθανότητες που μπορούν να εμφανιστούν σε μια περίπτωση και εξετάσαμε την απόδοσή του με το υπόλοιπο 20% του αρχείου καταγραφής συμβάντων.

Εξετάσαμε τα αποτελέσματα που πρόβλεψε το μοντέλο, με τα πραγματικά αποτελέσματα που εμπεριέχονταν στο 20% του δοκιμαστικού μέρους του αρχείου καταγραφής

συμβάντων και μπορούμε να δούμε αναλυτικά την πορεία του στον παρακάτω πίνακα, γνωστό και ως πίνακα σύγχυσης (confusion matrix – εικόνα 39).



Εικόνα 39. Confusion matrix.

Πίνακας 25. Ανάλυση πίνακα σύγχυσης (confusion matrix).

TN = 55, TP = 132, FN = 34, FP = 30	
accuracy score	0.745 ή 74,5%
precision score	0.814 ή 81,4%
recall score	0.795 ή 79,5%
f1 score	0.804 ή 80,4%

Ένας πίνακας σύγχυσης, χρησιμοποιείται για να περιγράψει την απόδοση ενός μοντέλου ταξινόμησης. Απαρτίζεται από τέσσερα κελιά, τα TP, TN, FP και FN. Μέσω του πίνακα αυτού, άμεσα παρατηρούνται τα εξής:

- υπάρχουν δύο προβλεπόμενες τιμές (0 και 1 ή g-pay compensation και h-reject request),
- συνολικά έγιναν 251 προβλέψεις,
- από τις 251 προβλέψεις, οι 162 αντιστοιχούν στη πρόβλεψη της δραστηριότητας h, ενώ οι υπόλοιπες 89 στη δραστηριότητα g και

- στην πραγματικότητα, οι 166 από τα 251 έγγραφες ανήκουν στη δραστηριότητα h, ενώ οι υπόλοιπες 85 στη g.

Για την κατανόηση της σημασίας των κελιών του πίνακα, TN – True negative, TP – True positive, FN – False negative και FP – false positive, η βασική ορολογία που τους περιγράφει φαίνεται παρακάτω ως εξής:

- TN: προβλέψαμε ότι η επόμενη δραστηριότητα ήταν 0 (g-pay compensation) και όντως ήταν 0. (Ανάλογο με την πρόβλεψη ότι ένας άντρας δεν είναι έγκυος και όντως δεν είναι έγκυος.)
- TP: προβλέψαμε ότι η επόμενη δραστηριότητα ήταν 1 (h-reject request) και όντως είναι 1. (Ανάλογο με την πρόβλεψη ότι μια γυναίκα είναι έγκυος και είναι έγκυος στην πραγματικότητα.)
- FN: προβλέψαμε ότι η επόμενη δραστηριότητα ήταν 0 (g-pay compensation) αλλά ήταν 1 (h-reject request). (Ανάλογο με την πρόβλεψη ότι μια γυναίκα δεν είναι έγκυος ενώ στην πραγματικότητα είναι.)
- FP: προβλέψαμε ότι η επόμενη δραστηριότητα ήταν 1 (h-reject request) αλλά ήταν 0 (g-pay compensation). (Ανάλογο με την πρόβλεψη ότι ένας άντρας είναι έγκυος ενώ δεν είναι.)

Πιο συγκεκριμένα, το μοντέλο μας έβγαλε 55 TN, 132 TP, 34 FN και 30 FP. Με γνώμονα τα παραπάνω, αξιολογούμε το μοντέλο με βάση τα:

- **Accuracy** = $\frac{TP+TN}{TP+TN+FP+FN}$, που περιγράφει πόσο συχνά είναι σωστό είναι το μοντέλο. Στην περίπτωσή μας ήταν 74.5%.
- **Precision** = $\frac{TP}{TP + FP}$, ακρίβεια, το ποσοστό το οποίο όταν το μοντέλο προβλέπει κάτι θετικά, πόσες φορές είναι σωστό. Στην περίπτωσή μας ήταν 81.4%.
- **Recall** = $\frac{TP}{TP + FN}$, ευαισθησία, το ποσοστό το οποίο όταν στην πραγματικότητα είναι θετικό, πόσο συχνά προβλέπει θετικά. Στην περίπτωσή μας ήταν 79.5%.

- **F1 score** = $2 * \frac{\text{Precision} * \text{Recall}}{\text{Precision} + \text{Recall}}$, συνολικό μέτρο της ακρίβειας ενός μοντέλου που συνδυάζει την ακρίβεια και την ανάκληση-ευαισθησία. Στην περίπτωση μας ήταν 80.4%.
- **Misclassification rate** = $\frac{\text{FP} + \text{FN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}}$, συνολικά πόσο συχνά είναι λάθος το μοντέλο. Στην περίπτωση μας ήταν 25.5%.

Το μοντέλο που δημιουργήθηκε για να προβλέψει την επόμενη επιχειρησιακή δραστηριότητα, εκπαιδεύτηκε πάνω σε 1004 εγγραφές του αρχείου καταγραφής γεγονότων (80% του συνολικού αρχείου καταγραφής). Έπειτα, αξιολογήθηκε με βάση τις 251 εγγραφές που ορίσαμε ως δεδομένα δοκιμής, δηλαδή συγκρίθηκαν τα αποτελέσματα που έδωσε το μοντέλο με τα πραγματικά δεδομένα που υπήρχαν στο αρχείο καταγραφής.

Στατιστικά που αφορούν την απόδοση του μοντέλου φάνηκαν στον πίνακα σύγκρισης, ένα εργαλείο γνωστό στο τομέα της μηχανικής μάθησης, που μέσω μαθηματικών τύπων αξιολογεί ένα μοντέλο.

Δεδομένου όμως όλων των παραπάνω μετρήσεων, τι κρατάμε τελικά για την απόδοση του μοντέλου; Αν και δεν υπάρχει μια τυπική απάντηση πάνω σε αυτό το κομμάτι, εξαρτάται ανάλογα την πτυχή του προβλήματος. Εάν η ακρίβεια των αποτελεσμάτων είναι το κέντρο του ενδιαφέροντος, τότε κρίνεται με βάση την ακρίβεια. Εάν μας ενδιαφέρει περισσότερο η ευαισθησία του μοντέλου βασιζόμαστε στην ευαισθησία κ.ο.κ. .

Στην περίπτωση που εξετάστηκε και με βάση τα παραπάνω αποτελέσματα, με ποσοστό λάθους πρόβλεψης στο 25,5%, παρατηρείται ότι η επόμενη δραστηριότητα που αναζητάμε, δεδομένου ενός «μοτίβου» δραστηριοτήτων που προηγούνται, προβλέπεται σωστά με ακρίβεια 7,5/10.

5. Συμπεράσματα και Μελλοντική Εργασία

Στην παρούσα διπλωματική εργασία προτάθηκε ένα σχέδιο με το οποίο εργαζόμαστε πάνω σε ένα αρχείο καταγραφής συμβάντων μιας επιχείρησης και μέσω πιθανοτικών μοντέλων μάθησης καταφέρνουμε να διαχειριστούμε την αβεβαιότητα που πολλές φορές επικρατεί κατά τις μεταβάσεις μεταξύ των δραστηριοτήτων μιας επιχείρησης.

Καθώς οι διαδικασίες αποτελούν το βασικότερο συστατικό μιας επιχείρησης, είναι χρήσιμη η ανάλυση και η συνεχής παρακολούθησή τους, προκειμένου να αναγνωρίζονται σημεία και καταστάσεις που διαφέρουν από την «κανονικότητα». Η αναπαράστασή τους σε ένα επιχειρησιακό μοντέλο καθιστά εύκολη την αναγνώριση της ροής μιας εργασίας ή διεργασιών, καθώς και των σχέσεων μεταξύ τους αλλά και τον προσδιορισμό σημείων συμφόρησης. Στις περιπτώσεις που δεν παρατηρείται ταύτιση ενός τέτοιου μοντέλου με τις πραγματικές διαδικασίες στην επιχείρηση, έρχεται η έννοια της αβεβαιότητας. Οι σχέσεις μεταξύ των διαδικασιών δεν είναι ξεκάθαρες και ορισμένες φορές δεν μπορούμε να γνωρίζουμε την πραγματική σειρά εκτέλεσης των διαδικασιών καθώς και αν όντως ολοκληρώνονται οι διαδικασίες.

Δημιουργήθηκε ένα μοντέλο που μπορεί και διαχειρίζεται την αβεβαιότητα, όσο περίπλοκες και να είναι οι εξαρτώμενες σχέσεις μεταξύ των διαδικασιών και προβλέπει μια επιχειρησιακή δραστηριότητα, δεδομένου του αρχείου καταγραφής συμβάντων. Το πιθανοτικό μοντέλο μάθησης που εφαρμόστηκε στην παρούσα διπλωματική ήταν τα μπεϋζιανά δίκτυα (Bayesian networks).

Συνοπτικά, ως πρώτο βήμα εφαρμόστηκαν τεχνικές εξόρυξης διεργασιών με στόχο την ανακάλυψη μοντέλων επιχειρησιακών διεργασιών. Έγινε μια εκτενής ανάλυση πάνω σε αυτά ώστε να πάρουμε μια ιδέα για τις διεργασίες που εμπεριέχονται στο αρχείο καταγραφής συμβάντων. Έπειτα, δημιουργήθηκε το μπεϋζιανό δίκτυο που μπορεί και επεξεργάζεται τα παραπάνω και έτσι προβαίνουμε στην επεξεργασία της αβεβαιότητας αλλά και στην πρόβλεψη πάνω στις επιχειρησιακές δραστηριότητες.

Η παρουσία μιας προσέγγισης επί του θέματος ήταν αναγκαία, καθώς κατάφερε και «έσπασε» τη διαδικασία αυτή σε μικρότερα μέρη καθώς και εξάλειψε όλα τα προβλήματα που θα εμφανιζόντουσαν στην πορεία.

Τέλος, έγιναν πειράματα πάνω σε ένα πραγματικό αρχείο καταγραφής συμβάντων μιας αεροπορικής εταιρείας και καταφέραμε και είδαμε έμπρακτα την προσέγγιση που αναφέραμε προηγουμένως.

Μερικά προβλήματα που παρουσιάστηκαν κατά την εκπόνηση της διπλωματικής εργασίας είναι:

- Αρχείο καταγραφής που εμπεριέχει κύκλους στις διαδικασίες του: Όπως αναλύθηκε, ένα μπεϋζιανό δίκτυο πρόκειται για ένα κατευθυνόμενο άκυκλο γράφημα. Η εμφάνιση κύκλων στο αρχείο καταγραφής δεν επιτρέπουν την «αυτόματη» μετάβαση ενός επιχειρησιακού μοντέλου σε ένα μπεϋζιανό δίκτυο. Χρειάζεται περεταίρω επεξεργασία για την μετατροπή αυτή και επομένως η ανάπτυξη μιας λογικής που θα μπορεί να το υλοποιεί. Η προσέγγιση που ακολουθήθηκε, προσαρμόστηκε γνωρίζοντας αυτό το πρόβλημα και αντιμετώπισε τους κύκλους.
- Μερικές τεχνικές αφαίρεσης κύκλων αφαιρούσαν σημαντικό μέρος των δεδομένων του αρχείου καταγραφής δεδομένων: Όπως προηγήθηκε, οι κύκλοι σε ένα αρχείο καταγραφής είναι ένα μείζον πρόβλημα που χρήζει αντιμετώπισης, καθώς ένα μπεϋζιανό δίκτυο είναι άκυκλο. Ένα πλήθος τεχνικών αφαίρεσης των κύκλων βασίζονται σε ευρετικούς αλγορίθμους (ένα από τα παραδείγματα είναι να κρατάμε την ακμή στους κόμβους με κύκλο με τις πιο πολλές διαδρομές) οι οποίοι ουσιαστικά αφαιρούν δεδομένα από το αρχείο καταγραφής γεγονότων, κάτι που δεν μας ωφελεί ιδιαίτερα καθώς, αργότερα, στη δημιουργία του πιθανοτικού μοντέλου θα χρειαστούμε όσο το δυνατόν περισσότερες εγγραφές για την εκπαίδευσή του. Οπότε και για τη λύση του προβλήματος αυτού, έγινε στροφή σε τεχνικές αφαίρεσης κύκλων διαφορετικής ιδεολογίας.

Τέλος, όσον αφορά μελλοντικές έρευνες και εργασίες, μερικά από τα θέματα που θα μπορούσαν να εξεταστούν για να εμπλουτιστεί η παρούσα διπλωματική εργασία είναι:

- Η βελτίωση του πιθανοτικού μοντέλου με στόχο την 100% επιτυχία στις προβλέψεις επιχειρησιακών δραστηριοτήτων,
- περισσότερες επιλογές για τη διαχείριση κύκλων κατά τη δημιουργία του άκυκλου γραφήματος που θα χρησιμοποιηθεί για το μπεϋζιανό δίκτυο, όπως τεχνικές

mutual information και εφαρμογή μηχανικής μάθησης και decision mining (Gini index, entropy, information gain) και

- τη δυνατότητα συμπλήρωσης ενός ελλιπούς αρχείου καταγραφής συμβάντων βάσει τη συμπεριφορά που έχει παρατηρηθεί.

Βιβλιογραφικές Αναφορές

Ελληνικές βιβλιογραφικές αναφορές

- [16] Παναγιώτου, Ν., Ευαγγελόπουλος, Ν., Κατημερτζόγλου, Π., & Γκαγιαλής, Σ. (2013). ΔΙΑΧΕΙΡΙΣΗ ΕΠΙΧΕΙΡΗΣΙΑΚΩΝ ΔΙΑΔΙΚΑΣΙΩΝ: ΟΡΓΑΝΩΣΗ, ΑΝΑΔΙΟΡΓΑΝΩΣΗ ΚΑΙ ΒΕΛΤΙΩΣΗ (pp. 23-31). Αθήνα: Εκδόσεις Κλειδάριθμος.
- [27] Συλλογικό, Έ. (2020). ΣΤΟΙΧΕΙΑ ΠΙΘΑΝΟΤΗΤΩΝ ΚΑΙ ΣΤΑΤΙΣΤΙΚΗΣ Γ ΛΥΚΕΙΟΥ. Αθήνα: ΙΤΥΕ - ΔΙΟΦΑΝΤΟΣ.

Ξενόγλωσσες βιβλιογραφικές αναφορές

- [1] Oxford, L., 2021. process. In: Oxford Languages. [online] Available at: <https://www.oxfordlearnersdictionaries.com/definition/english/process1_1?q=process> [Accessed 6 June 2021].
- [2] process. 2021. In Merriam-Webster.com. Retrieved Feb 22, 2021, from <https://www.merriam-webster.com/dictionary/process>
- [3] Process. (2021, February 22). In Wikipedia. <https://en.wikipedia.org/wiki/Process>
- [4] Weske, M. (2019). Business Process Management (3rd ed.). Berlin: Springer.
- [5] Von Rosing, M., Von Scheel, H., & Scheer, A. W. (2014). The Complete Business Process Handbook: Body of Knowledge from Process Modeling to BPM, Volume 1 (Vol. 1). Morgan Kaufmann.
- [6] Jeston, J. (2014). Business process management: practical guidelines to successful implementations. Routledge.
- [7] (2021, February 23) Definition of Business Process Management (BPM). Retrieved from Gartner, Inc. <https://www.gartner.com>

- [8] Palmer, N. (2021). What is BPM?. Retrieved 23 February 2021, from <https://bpm.com/what-is-bpm>
- [9] van der Aalst, W. (2016). *Process Mining: Data science in action* (2nd ed.). Berlin, Heidelberg: Springer.
- [10] Vasilecas, O., Savickas, T., & Lebedys, E. (2014, October). Directed acyclic graph extraction from event logs. In *International Conference on Information and Software Technologies* (pp. 172-181). Springer, Cham.
- [11] Wiecek, W., Bois, F. Y., & Gayraud, G. (2019). Structure learning of Bayesian networks involving cyclic structures. *arXiv preprint arXiv:1906.04992*.
- [12] Sutrisnowati, R. A., Bae, H., & Song, M. (2015). Bayesian network construction from event log for lateness analysis in port logistics. *Computers & Industrial Engineering*, 89, 53-66.
- [13] Dennis J.B. (2011) Petri Nets. In: Padua D. (eds) *Encyclopedia of Parallel Computing*. Springer, Boston, MA.
- [14] Munoz-Gama, J., & Carmona, J. (2010, September). A fresh look at precision in process conformance. In *International Conference on Business Process Management* (pp. 211-226). Springer, Berlin, Heidelberg.
- [15] Adriansyah, A., Munoz-Gama, J., Carmona, J., van Dongen, B. F., & van der Aalst, W. M. (2015). Measuring precision of modeled behavior. *Information systems and e-Business Management*, 13(1), 37-67.
- [17] Palmberg, K. (2009). *Beyond process management: Exploring organizational applications and complex adaptive systems* (Doctoral dissertation, Luleå tekniska universitet).
- [18] Dumas, M., La Rosa, M., Mendling, J., & Reijers, H. A. (2013). *Fundamentals of business process management* (Vol. 1, p. 2). Heidelberg: Springer.
- [19] van der Aalst W.M.P. (2009) Petri Nets. In: LIU L., ÖZSU M.T. (eds) *Encyclopedia of Database Systems*. Springer, Boston, MA.

- [20] van der Aalst W.M.P. (2009) Business Process Modeling Notation. In: LIU L., ÖZSU M.T. (eds) Encyclopedia of Database Systems. Springer, Boston, MA.
- [21] Buijs, J. C., van Dongen, B. F., & van der Aalst, W. M. (2014). Quality dimensions in process discovery: The importance of fitness, precision, generalization and simplicity. *International Journal of Cooperative Information Systems*, 23(01), 1440001.
- [22] Dakic, D., Stefanovic, D., Cosic, I., Lolic, T., & Medojevic, M. (2018). BUSINESS PROCESS MINING APPLICATION: A LITERATURE REVIEW. *Annals of DAAAM & Proceedings*, 29.
- [23] Van Der Aalst, W., Adriansyah, A., De Medeiros, A. K. A., Arcieri, F., Baier, T., Blickle, T., ... & Wynn, M. (2011, August). Process mining manifesto. In International Conference on Business Process Management (pp. 169-194). Springer, Berlin, Heidelberg.
- [24] Van Der Aalst, W. (2012). Process mining: Overview and opportunities. *ACM Transactions on Management Information Systems (TMIS)*, 3(2), 1-17.
- [25] Technology, F. (2021). PM4Py - Process Mining for Python. Retrieved 17 May 2021, from <https://pm4py.fit.fraunhofer.de/documentation#discovery>
- [26] Horný, M. (2014). Bayesian networks: A Technical report. *no*, 5, 15.
- [28] Ben-Gal, I. (2008). Bayesian networks. *Encyclopedia of statistics in quality and reliability*, 1.
- [29] University College Dublin. (2021). Statistics: Bayes' Theorem (worksheet). Belfield, Dublin, University College Dublin.
- [30] Bayes' Theorem. (2021). Retrieved 4 June 2021, from <https://www.mathsisfun.com/data/bayes-theorem.html>
- [31] Berti, A., van Zelst, S. J., & van der Aalst, W. (2019). Process mining for python (PM4Py): bridging the gap between process-and data science. *arXiv preprint arXiv:1905.06169*.

- [32] Ruggeri, F., Faltin, F., & Kenett, R. (2007). Bayesian networks. Encyclopedia of Statistics in quality and reliability.
- [33] pgmpy/pgmpy_notebook. (2021). Retrieved 4 June 2021, from https://github.com/pgmpy/pgmpy_notebook/blob/master/notebooks/2.%20Bayesian%20Networks.ipynb
- [34] Moreira, C. (2015). An Experiment on Using Bayesian Networks for Process Mining. *arXiv preprint arXiv:1503.07341*.

