



ΕΘΝΙΚΟ ΜΕΤΣΟΒΙΟ ΠΟΛΥΤΕΧΝΕΙΟ  
ΣΧΟΛΗ ΝΑΥΠΗΓΩΝ ΜΗΧΑΝΟΛΟΓΩΝ ΜΗΧΑΝΙΚΩΝ  
ΤΟΜΕΑΣ ΘΑΛΑΣΣΙΩΝ ΜΕΤΑΦΟΡΩΝ

Στατιστική Ανάλυση ναυτικών ατυχημάτων σε πλοία LNG LPG  
και ανάπτυξη αλγορίθμου εξόρυξης δεδομένων

ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

ΜΑΛΑΜΑΣ ΧΡΗΣΤΟΣ

Επιβλέπων : Νικόλαος Π. Βεντικός Αναπληρωτής Καθ. Ε.Μ.Π.

Τριμελής Επιτροπή :

κ. Ν.Π. Βεντικός Αναπληρωτής Καθηγητής Ε.Μ.Π.

κ. Α.Α. Γκίνης Αναπληρωτής Καθηγητής Ε.Μ.Π.

κ. Χ.Ι. Παπαδόπουλος Επίκουρος Καθηγητής Ε.Μ.Π.

Με επιφύλαξη κάθε δικαιώματος. All rights reserved

Απαγορεύεται η αντιγραφή, αποθήκευση ή διανομή της παρούσας εργασίας (ή τμήματος αυτής) για εμπορικούς σκοπούς. Επιτρέπεται η ανατύπωση, αποθήκευση και διανομή για σκοπούς μη κερδοσκοπικούς, εκπαιδευτικής ή ερευνητικής φύσης, υπό την προϋπόθεση της αναφοράς της πηγής προέλευσης και της διατήρησης του παρόντος μηνύματος. Ερωτήματα που αφορούν τη χρήση της εργασίας για κερδοσκοπικό σκοπό πρέπει να απευθύνονται προς τον συγγραφέα.

Οι απόψεις και τα συμπεράσματα που περιέχονται σε αυτή τη διπλωματική εργασία, εκφράζουν αποκλειστικά τον συγγραφέα και δεν πρέπει να ερμηνευθεί ότι αντιπροσωπεύουν τις επίσημες θέσεις του Εθνικού Μετσόβιου Πολυτεχνείου.

## ΕΥΧΑΡΙΣΤΙΕΣ

Από τη θέση αυτή θα ήθελα να ευχαριστήσω θερμά τον επιβλέποντα καθηγητή του τομέα Θαλασσίων Μεταφορών της Σχολής Ναυπηγών Μηχανολόγων Μηχανικών, Δρ. Νικόλαο Π. Βεντίκο για την ανάθεση και επίβλεψη της παρούσας διπλωματικής εργασίας. Η συνεργασία που είχαμε, η εμπιστοσύνη στο πρόσωπο μου και οι καίριες συμβουλές του ήταν στοιχεία απαραίτητα ώστε να έρθει εις πέρας η εργασία.

Θέλω να ευχαριστήσω και τον υποψήφιο διδάκτωρ Κωσταντίνο Λούζη για την πολύτιμη βοήθεια που μου παρείχε.

Επίσης θέλω να ευχαριστήσω την οικογένεια μου για την υπομονή και την στήριξη της όλα αυτά τα χρόνια και ιδιαιτέρως τους γονείς μου, για όσα μου έχουν προσφέρει.

Τέλος θα ήθελα να ευχαριστήσω τον συνάδελφο Αναστάσιο Βενέτη για την βοήθεια και τις σημειώσεις που παρείχε αφιλοκερδώς όλα αυτά τα χρόνια και οι οποίες συνέβαλαν στην περάτωση των σπουδών.

Μάλαμας Χρήστος

Αθήνα, Ιούλιος 2017

## ΠΕΡΙΛΗΨΗ

Η ασφάλεια των πλοίων στη θάλασσα αποτελούσε ανέκαθεν και θα συνεχίσει να αποτελεί ένα από τα σημαντικότερα μελετήματα όλων όσων ασχολούνται με τη ναυτιλία. Το ενδιαφέρον πολλαπλασιάζεται στην περίπτωση που το φορτίο που μεταφέρεται μπορεί να προκαλέσει οικολογικές καταστροφές.

Η παρούσα εργασία εστιάζει στο πρόβλημα των ναυτικών ατυχημάτων σε πλοία μεταφοράς LNG-LPG. Οι συνέπειες από ναυτικά ατυχήματα σε πλοία αυτής της κατηγορίας μπορούν να αποβούν καταστροφικές για το πλοίο, το πλήρωμα και το περιβάλλον. Έγινε λοιπόν μία προσπάθεια με χρήση της βάσης δεδομένων Seaweb να προσδιοριστούν οι παράγοντες οι οποίοι συμβάλουν στην ύπαρξη ναυτικών ατυχημάτων σε πλοία αυτής της κατηγορίας. Η μελέτη επικεντρώθηκε σε ατυχήματα που συνέβησαν μετά το 1996 ώστε τα εξαγόμενα αποτελέσματα να είναι συμβατά με τα πλοία που λειτουργούν σήμερα.

Αρχικά μελετήθηκαν τα κατασκευαστικά χαρακτηριστικά στοιχεία των υγραεριοφόρων πλοίων. Έγινε αναγνώριση των κινδύνων που εγκυμονούν σε αυτά και προς την κατεύθυνση αυτή πραγματοποιήθηκε στατιστική ανάλυση των ατυχημάτων βάσει των δεδομένων που εξήχθησαν από την βάση δεδομένων Seaweb ([http://www.sea-web.com/seaweb\\_welcome.aspx](http://www.sea-web.com/seaweb_welcome.aspx)). Βάσει αυτής της ανάλυσης παράχθηκαν διαγράμματα κατανομής ατυχημάτων και συχνοτήτων.

Τέλος έγινε μία προσπάθεια να εκπαιδευτεί ένας αλγόριθμος εξόρυξης δεδομένων με ικανότητες πρόβλεψης του τύπου ατυχήματος που είναι πιο πιθανόν να συμβεί σε ένα πλοίο με συγκεκριμένα χαρακτηριστικά.

## ABSTRACT

The safety of ships at sea has always been and will continue to be one of the most important issues for the whole shipping industry. Interest in this field is multiplied if the shipment carried can cause vast ecological disasters.

This paper focuses on the problem of maritime accidents on LNG-LPG Tankers. The consequences of maritime accidents on ships of this category can be devastating for the integrity of the ship, the crew and the environment. An effort was therefore made to use the Seaweb database ([http://www.sea-web.com/seaweb\\_welcome.aspx](http://www.sea-web.com/seaweb_welcome.aspx)) to identify the factors that contribute to the occurrence of maritime accidents on ships of this class. The study focused on accidents that occurred after 1996 so that the exported results are applicable on modern ships.

Firstly, the unique structural nature of LNG and LPG Tankers was researched. The most prominent dangers were identified and to this end a statistical analysis of accidents was carried out based on the data extracted from the Seaweb database. Based on this analysis, graphs depicting the distributions of accidents and frequencies were produced.

Finally, an attempt was made to train an algorithm which would be able to predict the most likely type of accident to happen on a ship given specific data.

## Περιεχόμενα

Κατάλογος εικόνων.....	8
Κατάλογος Πινάκων.....	9
Κατάλογος Σχημάτων.....	10
ΚΕΦΑΛΑΙΟ 1. ΕΙΣΑΓΩΓΗ.....	13
ΚΕΦΑΛΑΙΟ 2. Πλοία μεταφοράς Φυσικού Αερίου (LNG).....	15
Κεφάλαιο 2.1 Σχεδίαση δεξαμενών.....	18
Κεφάλαιο 2.2 Μέθοδοι μεταφοράς LNG.....	21
Κεφάλαιο 2.3 Στατιστικά στοιχεία σχετικά με τα πλοία μεταφοράς LNG & προβλήματα σχεδίασης.....	23
Κεφάλαιο 2.4 Κίνδυνοι και ατυχήματα σε LNG Tankers.....	25
Κεφάλαιο 2.4.1 Κίνδυνοι.....	25
Κεφάλαιο 2.4.2 Αξιολόγηση ρίσκου σε πλοία μεταφοράς LNG.....	31
ΚΕΦΑΛΑΙΟ 3. Αξιολόγηση και έλεγχος των δεδομένων με χρήση της μεθόδου $\chi^2$ .....	33
Κεφάλαιο 3.1 Απαραμετρική Συμπεριματολογία.....	33
Κεφάλαιο 3.2 Έλεγχος ανεξαρτησίας $\chi^2$ (Chi square Independence test).....	36
Κεφάλαιο 3.3 Επιλογή αξιολόγηση και ανάλυση Δεδομένων.....	39
Κεφάλαιο 3.3.1 Πληρότητα δεδομένων.....	40
Κεφάλαιο 3.3.2 Εξάρτηση είδους ατυχήματος – δεδομένων.....	41
ΚΕΦΑΛΑΙΟ 4. Στατιστική Ανάλυση Ατυχημάτων.....	51
Κεφάλαιο 4.1 Στατιστική ανάλυση ατυχημάτων υγραεριοφόρων πλοίων.....	53
Κεφάλαιο 4.1.1 Κατανομή στόλου.....	54
Κεφάλαιο 4.1.2 Συγκεντρωτική κατανομή ατυχημάτων.....	56
Κεφάλαιο 4.1.3 Collision.....	61
Κεφάλαιο 4.1.4 Wrecked/Stranded.....	66
Κεφάλαιο 4.1.5 Hull & Machinery (Mchy) Damage.....	71
Κεφάλαιο 4.1.6 Fire/Explosion.....	76
Κεφάλαιο 4.2 Συγκεντρωτικές συχνότητες ατυχημάτων.....	81
Κεφάλαιο 4.3 Συχνότητες ατυχημάτων συναρτήσει των ηλικιών.....	86
ΚΕΦΑΛΑΙΟ 5. Εξόρυξη Δεδομένων.....	91
Κεφάλαιο 5.1 Στάδια της Διαδικασίας Ανακάλυψης Γνώσης.....	93
Κεφάλαιο 5.2 Εξόρυξη Δεδομένων Εργασίες και Αλγόριθμοι.....	97

Στατιστική Ανάλυση Ναυτικών Ατυχημάτων σε πλοία LNG-LPG  
και Ανάπτυξη Αλγορίθμου Εξόρυξης Δεδομένων

Κεφάλαιο 5.2.1 Ορισμός .....	97
Κεφάλαιο 5.2.2 Μηχανισμοί Εξόρυξης Δεδομένων .....	99
Κεφάλαιο 5.2.3 Κατηγορίες Αλγορίθμων Εξόρυξης Δεδομένων .....	101
Κεφάλαιο 5.3 Εκπαίδευση αλγορίθμου εξόρυξης δεδομένων για ατυχήματα LNG-LPG πλοίων .....	113
5.3.1 Μέθοδος εκπαίδευσης και επιλογής αλγορίθμου .....	114
5.3.2 Εκπαίδευση και επιλογή αλγορίθμου για τα ατυχήματα LNG-LPG πλοίων .....	118
Κεφάλαιο 5.4 Εφαρμογή και αξιολόγηση τελικού αλγορίθμου .....	120
ΚΕΦΑΛΑΙΟ 6. ΣΥΜΠΕΡΑΣΜΑΤΑ ΚΑΙ ΠΡΟΤΑΣΕΙΣ ΓΙΑ ΠΕΡΑΙΤΕΡΩ ΜΕΛΕΤΗ .....	123
ΒΙΒΛΙΟΓΡΑΦΙΑ .....	127
ΠΑΡΑΡΤΗΜΑ Α – ΚΩΔΙΚΑΣ VBA .....	131

## Κατάλογος εικόνων

Εικόνα 1	Πρισματικές δεξαμενές φορτίου (Πηγή <a href="http://www.liquefiedgascarrier.com">http://www.liquefiedgascarrier.com</a> ) .....	17
Εικόνα 2	Σφαιρικές Δεξαμενές φορτίου (Πηγή <a href="http://www.liquefiedgascarrier.com">http://www.liquefiedgascarrier.com</a> ) .....	17
Εικόνα 3	Moss Tank της εταιρείας Moss Maritime (Πηγή <a href="http://www.slideshare.net/">www.slideshare.net/</a> ) .....	18
Εικόνα 4	Πρισματικού τύπου δεξαμενή IHI (Πηγή <a href="http://www.lngworldnews.com">http://www.lngworldnews.com</a> )	19
Εικόνα 5	Δομή δεξαμενής τύπου Mark III (Πηγή <a href="http://hhpindsight.com">http://hhpindsight.com</a> ) .....	19
Εικόνα 6	Το σύστημα μεμβρανών NO96 για LNGC (Πηγή <a href="http://www.gtt.fr/">http://www.gtt.fr/</a> ) .....	20
Εικόνα 7	Στάδια της διαδικασίας Ανακάλυψης γνώσης (Πηγή <a href="http://slideplayer.gr">http://slideplayer.gr</a> )	96
Εικόνα 8	Διαμερίσεις (αριστερά) και δομή δέντρων απόφασης (δεξιά) για ένα πρότυπο δέντρο κατηγοριοποίησης με τρεις κατηγορίες που ονομάζουμε 1, 2, και 3. Σε κάθε ένα ενδιάμεσο κόμβο, μια περίπτωση πηγαίνει στον αριστερό κόμβο απογόνων εάν και μόνο εάν η συνθήκη ικανοποιείται. Η προβλεπόμενη κατηγορία δίνεται κάτω από κάθε φυλλοειδή κόμβο (Πηγή <a href="http://slideplayer.gr">http://slideplayer.gr</a> ) .....	102
Εικόνα 9	Αναπαράσταση του βέλτιστου υπερεπίπεδου SVC για την περίπτωση γραμμικά διαχωριζόμενου χώρου. (Πηγή <a href="http://slideplayer.gr">http://slideplayer.gr</a> ) .....	105
Εικόνα 10	Μηχανισμός λειτουργίας κατηγοριοποιητών συνόλων (Πηγή <a href="http://slideplayer.gr">http://slideplayer.gr</a> ) .....	108
Εικόνα 11	Πρώτο βήμα εκπαίδευσης αλγορίθμου .....	114
Εικόνα 12	Δεύτερο βήμα εκπαίδευσης αλγορίθμου .....	114
Εικόνα 13	Τρίτο βήμα εκπαίδευσης αλγορίθμου .....	115
Εικόνα 14	Τέταρτο βήμα εκπαίδευσης αλγορίθμου .....	115
Εικόνα 15	Πέμπτο βήμα εκπαίδευσης αλγορίθμου .....	116
Εικόνα 16	Έκτο βήμα εκπαίδευσης αλγορίθμου .....	117
Εικόνα 17	Έβδομο βήμα εκπαίδευσης αλγορίθμου .....	117



## Κατάλογος Πινάκων

Πίνακας 1 Αριθμός ατυχημάτων συνολικά και μετά το 1996 .....	40
Πίνακας 2 Ελλειψές εγγραφές αριθμητικά και ποσοστιαία.....	40
Πίνακας 3 Κατανομή ατυχημάτων συναρτήσει του DWT .....	42
Πίνακας 4 Κατανομή ατυχημάτων συναρτήσει του Είδους του πλοίου.....	43
Πίνακας 5 Κατανομή ατυχημάτων συναρτήσει της σημαίας του πλοίου .....	44
Πίνακας 6 Κατανομή ατυχημάτων συναρτήσει της ζώνης ατυχήματος .....	45
Πίνακας 7 Κατανομή ατυχημάτων συναρτήσει της χώρας που κτίστηκε το πλοίο ....	47
Πίνακας 8 Κατανομή ατυχημάτων συναρτήσει της ηλικίας του πλοίου .....	48
Πίνακας 9 Συνολικές τιμές Ρi και Χ <sup>2</sup> για όλα τα δεδομένα.....	49
Πίνακας 10 Κατανομή ατυχημάτων ανά κατηγορία ατυχήματος.....	53
Πίνακας 11 Κατανομή ατυχημάτων μετά το 1996 .....	81
Πίνακας 12 Seaweb Fleet at Risk & Frequencies .....	83
Πίνακας 13 Αθροιστικό Fleet at Risk ανά πενταετία .....	86
Πίνακας 14 Κατανομή ατυχημάτων ανά πενταετία συναρτήσει της ηλικίας.....	86
Πίνακας 15 Σύγκριση αλγορίθμων δέντρων παλινδρόμησης (με tick δηλώνεται ότι υποστηρίζεται το χαρακτηριστικό).....	103
Πίνακας 16 Αποτελέσματα δοκιμών εκπαίδευσης αλγορίθμου.....	119

## Κατάλογος Σχημάτων

Σχήμα 1 Κατανομή ατυχημάτων συναρτήσει του DWT .....	42
Σχήμα 2 Κατανομή ατυχημάτων συναρτήσει του Είδους του πλοίου .....	43
Σχήμα 3 Κατανομή ατυχημάτων συναρτήσει της σημαίας του πλοίου .....	44
Σχήμα 4 Κατανομή ατυχημάτων συναρτήσει της ζώνης ατυχήματος .....	46
Σχήμα 5 Κατανομή ατυχημάτων συναρτήσει της χώρας που κτίστηκε το πλοίο .....	47
Σχήμα 6 Κατανομή ατυχημάτων συναρτήσει της ηλικίας του πλοίου .....	48
Σχήμα 7 Κατανομή του στόλου συναρτήσει του Ship Type .....	54
Σχήμα 8 Κατανομή του στόλου συναρτήσει της σημαίας.....	54
Σχήμα 9 Κατανομή του στόλου συναρτήσει του Country Built.....	55
Σχήμα 10 Συγκεντρωτική κατανομή ατυχημάτων συναρτήσει του Ship Type .....	56
Σχήμα 11 Συγκεντρωτική κατανομή ατυχημάτων συναρτήσει του Flag.....	57
Σχήμα 12 Συγκεντρωτική κατανομή ατυχημάτων συναρτήσει του Zone .....	58
Σχήμα 13 Συγκεντρωτική κατανομή ατυχημάτων συναρτήσει του Country Built.....	59
Σχήμα 14 Συγκεντρωτική κατανομή ατυχημάτων συναρτήσει του Age of Ship.....	60
Σχήμα 15 Κατανομή ατυχημάτων Collision συναρτήσει του Ship Type.....	61
Σχήμα 16 Κατανομή ατυχημάτων Collision συναρτήσει του Flag.....	62
Σχήμα 17 Κατανομή ατυχημάτων Collision συναρτήσει του Zone .....	63
Σχήμα 18 Κατανομή ατυχημάτων Collision συναρτήσει του Country Built .....	64
Σχήμα 19 Κατανομή ατυχημάτων Collision συναρτήσει του Age of Ship .....	65
Σχήμα 20 Κατανομή ατυχημάτων Wrecked/Stranded συναρτήσει του Ship Type.....	66
Σχήμα 21 Κατανομή ατυχημάτων Wrecked/Stranded συναρτήσει του Flag.....	67
Σχήμα 22 Κατανομή ατυχημάτων Wrecked/Stranded συναρτήσει του Zone .....	68
Σχήμα 23 Κατανομή ατυχημάτων Wrecked/Stranded συναρτήσει του Country Built .....	69
Σχήμα 24 Κατανομή ατυχημάτων Wrecked/Stranded συναρτήσει του Age of Ship ..	70
Σχήμα 25 Κατανομή ατυχημάτων Hull&Mchy Damage συναρτήσει του Ship Type ...	71
Σχήμα 26 Κατανομή ατυχημάτων Hull&Mchy Damage συναρτήσει του Flag .....	72
Σχήμα 27 Κατανομή ατυχημάτων Hull&Mchy Damage συναρτήσει του Zone.....	73
Σχήμα 28 Κατανομή ατυχημάτων Hull&Mchy Damage συναρτήσει του Country Built .....	74
Σχήμα 29 Κατανομή ατυχημάτων Hull&Mchy Damage συναρτήσει του Age of Ship	75
Σχήμα 30 Κατανομή ατυχημάτων Fire&Explosion συναρτήσει του Ship Type .....	76
Σχήμα 31 Κατανομή ατυχημάτων Fire&Explosion συναρτήσει του Flag .....	77
Σχήμα 32 Κατανομή ατυχημάτων Fire&Explosion συναρτήσει του Zone.....	78
Σχήμα 33 Κατανομή ατυχημάτων Fire&Explosion συναρτήσει του Country Built .....	79
Σχήμα 34 Κατανομή ατυχημάτων Fire&Explosion συναρτήσει του Age of Ship.....	80
Σχήμα 35 Κατανομή ατυχημάτων μετά το 1996 .....	82
Σχήμα 36 Σύγκριση Fleet at Risk Seaweb-Clarkson .....	84
Σχήμα 37 Συχνότητες ατυχημάτων με χρήση Seaweb και Clarkson .....	84

Σχήμα 38 Συχνότητες ατυχημάτων συναρτήσει της ηλικίας του πλοίου ανά πενταετία .....	87
Σχήμα 39 Συχνότητες ατυχημάτων συναρτήσει της χρονολογίας ανά ηλικιακή κατηγορία .....	87
Σχήμα 40 Κατανομή ατυχημάτων ανά τύπο ατυχήματος για ηλικίες πλοίων 25-34 έτη .....	89
Σχήμα 41 Σύγκριση κατανομής ατυχημάτων μεταξύ δεδομένων εκπαίδευσης και πιθανών εισαγωγών αλγορίθμου .....	121

Η σελίδα αυτή είναι σκόπιμα κενή.

## ΚΕΦΑΛΑΙΟ 1. ΕΙΣΑΓΩΓΗ

Από τις πρώτες κοινωνίες που αναπτύχθηκαν έγινε άμεσα αντιληπτή η ανάγκη για μεταφορά εμπορευμάτων και ιδεών. Συνεπώς ναυπηγήθηκαν τα πρώτα πλοία για να ικανοποιηθεί αυτή η ανάγκη μέσω θαλασσίων οδών. Με την ναυπήγηση των πρώτων πλοίων εμφανίστηκαν και τα πρώτα ναυτικά ατυχήματα.

Ως Ναυτικό ή Θαλάσσιο ατύχημα χαρακτηρίζεται γενικά κάθε θαλάσσιο συμβάν που έχει ως συνέπεια την απώλεια ή βλάβη πλοίου ή φορτίου του ή τραυματισμό/απώλεια ανθρώπινης ζωής [54]. Το ναυτικό ατύχημα παρουσιάζεται και εξετάζεται ανάλογα με τα αίτια που το προκαλούν, τις συνθήκες κάτω από τις οποίες επήλθε και τις συνέπειες που προκάλεσε.

Στην παρούσα εργασία εξετάζονται τα πλοία μεταφοράς LNG-LPG τα οποία πρωτοεμφανίστηκαν την δεκαετία του 1960 στον τομέα της ναυτιλίας. Αν και αποτελούν μόνο το 9% του παγκόσμιου στόλου [54] οι προφυλάξεις που λαμβάνονται σε αυτά είναι ιδιαίτερα μεγάλες σε σχέση με τα περισσότερα από τα υπόλοιπα είδη πλοίων καθώς μεταφέρουν εύφλεκτες ουσίες και σοβαρά ατυχήματα σε αυτά τα πλοία θα μπορούσαν να αποβούν καταστροφικά.

Για το λόγο αυτό η κατασκευή αυτών των πλοίων είναι ιδιαίτερη περίπλοκη και περιλαμβάνει πολλές δικλείδες ασφαλείας για την αποφυγή διαρροής φορτίου προς το περιβάλλον. Στην συνέχεια περιγράφεται αναλυτικά η κατασκευή πλοίων LNG τα οποία έχουν το μεγαλύτερο ενδιαφέρον, την μεγαλύτερη περιπλοκότητα και τους πιο αυστηρούς κανονισμούς εκ των δύο τύπων πλοίων που θα μελετηθούν.

Ένα ατύχημα σε πλοίο LNG-LPG μπορεί να έχει σοβαρές συνέπειες τόσο στο ίδιο το πλοίο και το πλήρωμα όσο και στο περιβάλλον. Όπως προκύπτει από τη στατιστική ανάλυση που έγινε στα πλαίσια αυτής της εργασίας, τα συχνότερα εμφανιζόμενα ατυχήματα είναι ατυχήματα σύγκρουσης ή επαφής πλοίων, προσάραξης πλοίων ή επαφής τους με σταθερά αντικείμενα, ατυχήματα έκρηξης φωτιάς ή και αστοχίας μηχανικών εξαρτημάτων και της γάστρας. Τέτοια ατυχήματα μπορούν να οδηγήσουν με μαθηματική βεβαιότητα στην απώλεια ευστάθειας και την ολική απώλεια του πλοίου.

Συνεπώς κρίνεται αναγκαίο να προσδιοριστούν τα αίτια που οδηγούν σε κάθε ένα είδος ατυχήματος. Ένας τρόπος για να διεξαχθεί αυτή η ανάλυση είναι η μελέτη των ατυχημάτων που έχουν ήδη συμβεί. Κατά κύριο λόγο η νομοθεσία και οι κανονισμοί της ναυτιλίας παγκοσμίως μεταβάλλονται κάθε φορά που συμβαίνει ένα ατύχημα μεγάλου μεγέθους. Απόδειξη αυτού είναι το γεγονός ότι η SOLAS θεσπίστηκε μετά το ατύχημα του Τιτανικού το 1912.

## Στατιστική Ανάλυση Ναυτικών Ατυχημάτων σε πλοία LNG-LPG και Ανάπτυξη Αλγορίθμου Εξόρυξης Δεδομένων

Στην παρούσα εργασία θα γίνει μία προσπάθεια να επεξεργαστούμε τα ατυχήματα συνολικά και να εξάγουμε συμπεράσματα γενικής χρήσης για όλα τα πλοία μεταφοράς LNG-LPG. Αυτό το οποίο θα πραγματοποιηθεί είναι μία στατιστική ανάλυση ατυχημάτων βάσει των στοιχείων που έχουμε στην διάθεση μας. Τέλος θα γίνει και η προσπάθεια ανάπτυξης ενός αλγορίθμου με χρήση τεχνικών εξόρυξης δεδομένων για την πρόβλεψη του τύπου τυχόν ατυχημάτων στο μέλλον.

## ΚΕΦΑΛΑΙΟ 2. Πλοία μεταφοράς Φυσικού Αερίου (LNG)

Το φυσικό αέριο, ένα καθαρό, μη τοξικό καύσιμο, είναι τώρα η τρίτη σπουδαιότερη πηγή ενέργειας στον κόσμο [44], μετά το πετρέλαιο και τον άνθρακα. Επειδή το αέριο σε υγροποιημένη μορφή καταλαμβάνει πολύ λιγότερο όγκο, και λόγω της κρίσιμης θερμοκρασίας του υγροποιημένου μεθανίου, (-82.6 °C) η θαλάσσια μεταφορά του έχει μόνο νόημα από εμπορική άποψη εάν μεταφέρεται σε υγροποιημένη μορφή υπό ατμοσφαιρική πίεση και έτσι δημιουργεί μια μεγάλη πρόκληση στους μηχανικούς, κυρίως επειδή πρέπει να μεταφερθεί σε πολύ χαμηλή θερμοκρασία. Τα λεγόμενα Υγραεριοφόρα πλοία είναι ειδικός τύπος εμπορικών πλοίων, δεξαμενόπλοια που μεταφέρουν υγροποιημένα αέρια.

Στις σύγχρονες θαλάσσιες μεταφορές παρατηρούνται δύο βασικοί τύποι υγραεριοφόρων πλοίων:

1. Τα Υγραεριοφόρα φυσικού αερίου, λεγόμενα και δεξαμενόπλοια ή τάνκερ φυσικού αερίου, (Gas Tanker-LNG), εκ του Liquefied Natural Gas, και
2. Τα Υγραεριοφόρα αερίου πετρελαίου, λεγόμενα αντίστοιχα δεξαμενόπλοια ή τάνκερ πετρελαϊκού αερίου (Gas Tankers-LPG), εκ του Liquefied Petroleum Gas. Υγροποιημένο αέριο πετρελαίου είναι το γενικό όνομα που δίνεται για το προπάνιο, το βουτάνιο και τα μίγματα των δύο. Αυτά τα προϊόντα μπορούν να ληφθούν από τον καθαρισμό του ακατέργαστου πετρελαίου. Όταν παράγονται κατά αυτόν τον τρόπο κατασκευάζονται συνήθως υπό σταθερή ατμοσφαιρική πίεση.

Η ιδέα της μεταφοράς του φυσικού αερίου σε υγρή κατάσταση κατοχυρώθηκε αρχικά το Μάιο του 1915 με το δίπλωμα ευρεσιτεχνίας από τον Godfrey L. Cabot. Η προσπάθεια να πραγματοποιηθεί αυτή η ιδέα ήρθε ξανά στην επιφάνεια το 1952 από τον Willard L. Morrison, και ακολούθως από τον J.J. Henry το 1954 όταν ανέπτυξαν τις φορτηγίδες μεταφοράς υγροποιημένου μεθανίου, καλούμενες τότε "μεθανιοφόρα", και έφεραν ορθογώνιες τραπεζοειδείς δεξαμενές (Ffooks, 1993). Εντούτοις η ιδέα της μεταφοράς υγροποιημένου αερίου έγινε πραγματικότητα με το πρώτο πλοίο μεταφοράς LNG, που ονομάστηκε Suehiro Maru No. 8, με συνολική μεταφορική ικανότητα 150 m<sup>3</sup> LNG, που χρησιμοποιήθηκε από το 1962 έως το 1983 χρησιμοποιώντας μια μηχανή εσωτερικής καύσης σαν κύριο μέσο πρόωσης. Το πρώτο πλοίο μεταφοράς LNG που λειτουργούσε με χρήση ατμοστροβίλων με ονομασία Methane Princess (5034 τόνοι dwt) πραγματοποίησε το παρθενικό του ταξίδι από το Calcasieu στην ακτή του Κόλπου της Λουϊζιάνας στις 25-01-1959, αερίου και έπλευσε προς το Ηνωμένο Βασίλειο, ενώ το δεύτερο πλοίο μεταφοράς LNG, λειτούργησε από το 1964 έως το 1998 με μια μεταφορική ικανότητα της τάξης

των 27.400 m<sup>3</sup> (MAN, 2009). Από τότε οι αριθμοί, οι διαμορφώσεις και τα μεγέθη των πλοίων μεταφοράς LNG έχουν συνεχίσει να αυξάνονται.

Ένα πλοίο μεταφοράς LNG (LNGC - LNG Carrier), χρησιμοποιείται για να μεταφέρει το υγροποιημένο φυσικό αέριο σε μια κατάσταση κορεσμού σε ατμοσφαιρική πίεση χωρίς να υπάρχει η απαίτηση μιας δεξαμενής πίεσης (Oka et Al, 2004). Λόγω της φύσης του LNG, που πρέπει να υπόκειται σε πολύ χαμηλή θερμοκρασία (-160 °C) για να διατηρήσει την υγρή του κατάσταση, το σχέδιο και η κατασκευή αυτού του τύπου σκάφους γίνονται πολύ περίπλοκα. Επιπλέον, κάθε κύριο κατασκευαστικό block ενός πλοίου μεταφοράς LNG συσχετίζεται με τα άλλα, αυξάνοντας κατά συνέπεια την πολυπλοκότητα της κατασκευής του σκάφους λόγω της ύπαρξης εκτεταμένου συστήματος σωληνώσεων για την ψύξη του φορτίου και την διατήρηση της πίεσης και της θερμοκρασίας μέσα στις δεξαμενές. Αυτό καθιστά τα πλοία μεταφοράς LNG να είναι τα πιο ακριβά, σύνθετα και ενδεχομένως επικίνδυνα πλοία μεταφοράς φορτίου που λειτουργούν στους ωκεανούς σε παγκόσμια κλίμακα [54]. Υποθέτοντας ότι χρησιμοποιείται ένα παρόμοιο σύστημα συγκράτησης, καθώς το μέγεθος του σκάφους αυξάνεται, το ποσό του εξατμιζόμενου αερίου (Boil-off Gas, συντ. BOG) που παράγεται θα αυξηθεί επίσης. Το BOG χρειάζεται να αφαιρεθεί προκειμένου να αποτραπεί η συγκέντρωση πίεσης στις δεξαμενές. Αυτό μπορεί να χρησιμοποιηθεί ως καύσιμο για την πρόωση του πλοίου (υβριδικά σε συνδυασμό με τον κινητήρα Diesel) ή το μπορεί να μετατραπεί πάλι σε υγροποιημένη κατάσταση και να επιστραφεί στη δεξαμενή για να διατηρηθεί το αρχικό επίπεδο πλήρωσης της δεξαμενής. Αυτή η τελευταία επιλογή απαιτεί επιπρόσθετη ισχύ και μέγεθος των εγκαταστάσεων επανυγροποίησης προκειμένου να ρυθμιστεί ο όγκος του BOG που έχει παραχθεί. Ταυτόχρονα, απαιτείται επαρκής δύναμη πρόωσης για να υπερνικήσει τη συνολική υδροδυναμική αντίσταση και αυτή των κυματισμών που παράγονται από τις αλλαγές στη μορφή της γάστρας και της εξωτερικής βρεχόμενη επιφάνειας λόγω της πρόσθεσης 2 ακόμα διαχωριστικών τοιχωμάτων μεταξύ φορτίου και θάλασσας.

Το LNG αποθηκεύεται μέσα σε συστήματα θερμικής συγκράτησης, που η θερμοκρασία διατηρείται στο απαιτούμενο επίπεδο η οποία είναι αναγκαία για να αποτραπεί ο βρασμός στη θερμοκρασία κορεσμού του σε όλο το ταξίδι. Η διείσδυση της εξωτερικής θερμότητας από το περιβάλλον (αέρα και θάλασσα) στο φορτίο μέσω του συστήματος συγκράτησής, μαζί με την επίδραση των μηχανικών επιταχύνσεων ως αποτέλεσμα των έξι βαθμών ελευθερίας του σκάφους κίνησης ελευθερίας στα κύματα, και οι γενικές διαδικασίες φορτοεκφόρτωσης, υποβοηθούν στην εξάτμιση του LNG (Ohira et Al, 2002). Δεδομένου ότι το εξατμισμένο αέριο παράγεται σε όλο το ταξίδι, η συνεχής αφαίρεση αυτού του αερίου απαιτείται προκειμένου να αποτραπεί μια αύξηση της πίεσης μέσα στη δεξαμενή φορτίου λόγω της αυξανόμενης λανθάνουσας ενέργειας θερμότητας. Η διαδικασία της



αφαίρεσης της λανθάνουσας ενέργειας θερμότητας ψύχει εκ' νέου την υπόλοιπη ποσότητα LNG.

Αυτήν την περίοδο υπάρχουν δύο πιο ευρέως διαδεδομένοι τύποι σχεδίων για δεξαμενές φορτίου LNG οι πρισματικές (prismatic) τύπου B (IHI) και οι σφαιρικές δεξαμενές των οποίων τα άνω τμήματα υπερέχουν του κυρίου καταστρώματος του πλοίου της εταιρείας Maritime. Ακόμα πιο σύγχρονα υγραεριοφόρα φέρουν δεξαμενές τύπου μεμβράνης. Οι 2 πιο διαδεδομένοι τύποι δεξαμενών φορτίου LNG παρουσιάζονται στην παρακάτω εικόνα.



Εικόνα 1 Πρισματικές δεξαμενές φορτίου (Πηγή <http://www.liquefiedgascarrier.com>)



Εικόνα 2 Σφαιρικές Δεξαμενές φορτίου (Πηγή <http://www.liquefiedgascarrier.com>)

## Κεφάλαιο 2.1 Σχεδίαση δεξαμενών

Όπως προαναφέραμε υπάρχουν δύο είδη δεξαμενών όπου μπορούμε να τις κατηγοριοποιήσουμε σύμφωνα με τον τρόπο κατασκευής τους σε:

1. Αυτοστηριζόμενες (self-supporting)
2. Μembrανικού τύπου (membrane-type)

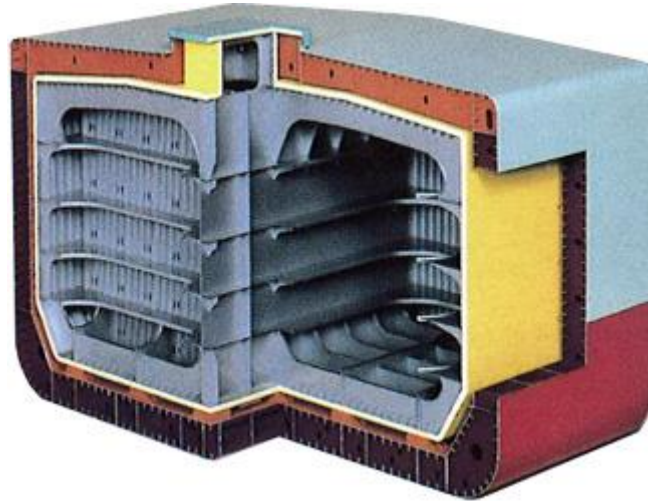
Έχουν αναπτυχθεί και σχεδιαστεί και διάφορες υποκατηγορίες των δύο παραπάνω τύπων LNG δεξαμενών με τις βασικότερες από αυτές να είναι:

**Moss tanks:** Αυτό το σχέδιο ανήκει στη νορβηγική Ναυτιλιακή Εταιρεία Moss Maritime και είναι μια σφαιρική δεξαμενή. Τα περισσότερα σκάφη τύπου Moss έχουν 4 ή 5 δεξαμενές. Το εξωτερικό της δεξαμενής έχει ένα παχύ στρώμα από μονωτικό αφρό που είτε τοποθετείται στο πάνελ ή σε πιο μοντέρνα σχέδια τυλίγεται στη δεξαμενή. Πάνω σε αυτή τη μόνωση υπάρχει ένα λεπτό στρώμα που επιτρέπει τη μόνωση. Αυτές οι δεξαμενές πρέπει να διατηρούνται στεγνές, με μια ατμόσφαιρα αζώτου. Αυτή η ατμόσφαιρα ελέγχεται συνεχώς για μεθάνιο, το οποίο θα αναφέρει μια διαρροή της δεξαμενής. Επίσης, το εξωτερικό της δεξαμενής ελέγχεται τακτικά σε διάστημα περίπου 3 μηνών για κρύα σημεία που θα υποδεικνύουν βλάβη στη μόνωση. Η δεξαμενή στηρίζεται γύρω από την περιφέρεια της στο ύψος του ισημερινού από ένα δακτυλίου ο οποίος εδράζεται σε μια μεγάλη κυκλική βάση που μεταφέρει το βάρος της δεξαμενής στη γάστρα του πλοίου. Αυτή η βάση επιτρέπει στη δεξαμενή να διαστέλλεται και συστέλλεται κατά τη διάρκεια της απόψυξης και της αναθέρμανσης καθώς μπορεί να συστέλλεται ή να διαστέλλεται περίπου 60cm. Λόγω αυτής της διαστολής και συστολής όλες οι σωληνώσεις στην δεξαμενή διέρχονται από το ανώτερο σημείο και συνδέονται με τις σωληνώσεις του πλοίου μέσω ευέλικτων συνδέσμων.



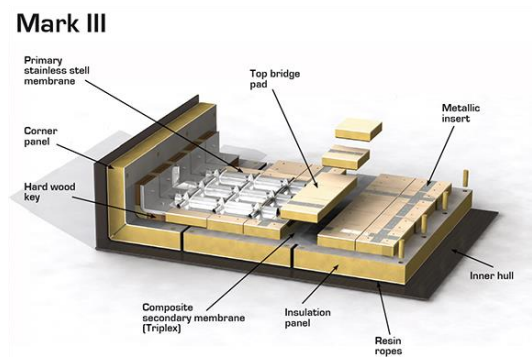
Εικόνα 3 Moss Tank της εταιρείας Moss Maritime (Πηγή [www.slideshare.net/](http://www.slideshare.net/))

**IHI:** Δεξαμενή που αναπτύχθηκε από την Ishikawajima-Harima Heavy Industries. Πρόκειται για αυτοφερόμενη πρισματική δεξαμενή τύπου Β από αλουμίνιο. Οι δεξαμενές τύπου Β περιορίζουν τα προβλήματα που προκαλούνται από τα χτυπήματα σε σχέση με τις δεξαμενές με μεμβράνες οι οποίες μπορεί να σπάσουν λόγω της πρόσκρουσης και να καταστρέψουν τη γάστρα του πλοίου.



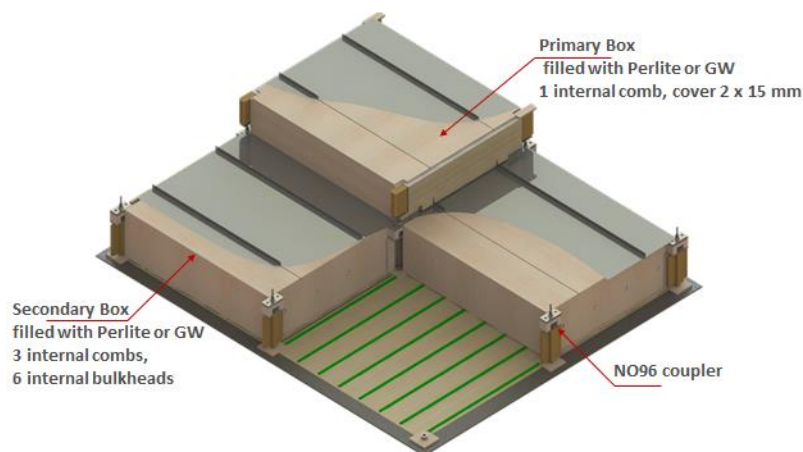
Εικόνα 4 Πρισματικού τύπου δεξαμενή IHI (Πηγή <http://www.lngworldnews.com>)

**TGZ Mark III:** Αυτός ο σχεδιασμός αναπτύχθηκε αρχικά από την Technigaz και είναι μια δεξαμενή τύπου μεμβράνης. Η μεμβράνη εδώ αποτελείται από ανοξείδωτο χάλυβα με κυματοειδή μορφή για να βοηθήσει στην απορρόφηση της θερμικής συστολής όταν η δεξαμενή έχει κρυώσει. Το κυρίως επίστρωμα, είναι κατασκευασμένα από ανοξείδωτο χάλυβα κυματοειδές με περίπου 1,2 cm πάχος είναι το ένα σε άμεση επαφή με το υγρό φορτίο (ή με τους ατμούς σε περίπτωση άδειας δεξαμενής). Αυτό ακολουθείται από μια βασική μόνωση η οποία με τη σειρά της καλύπτεται από ένα δευτερεύον επίστρωμα που έχει κατασκευαστεί από ένα υλικό που ονομάζεται "triplex" και το οποίο είναι ουσιαστικά ένα φύλλο μετάλλου ανάμεσα σε φύλλα υαλοβάμβακα συμπιεσμένα μαζί.



Εικόνα 5 Δομή δεξαμενής τύπου Mark III (Πηγή <http://hhpindsight.com>)

**NO96:** Τύπος δεξαμενής που αναπτύχθηκε από τη GazTransport. Πρόκειται για δεξαμενή τύπου μεμβράνης που αποτελείται από μία πρωτεύουσα και μια δευτερεύουσα λεπτή μεμβράνη από το υλικό Invar το οποίο έχει μηδενικές θερμικές συστολές. Η μόνωση είναι φτιαγμένη από ειδικά κουτιά από κόντρα πλακέ γεμισμένα με περλίτη που συνεχώς διαβρέχεται από αέριο Άζωτο. Η δομική ακεραιότητα και των δύο μεμβρανών ελέγχεται διαρκώς από την ανίχνευση υδρογονανθράκων στο άζωτο.



Εικόνα 6 Το σύστημα μεμβρανών NO96 για LNGC (Πηγή <http://www.gtt.fr/>)

**CS1:** Πρόκειται για μια δεξαμενή που σχεδιάστηκε μετά από τη συγχώνευση των εταιριών Technigaz και GazTransport και αποτελείται από τα καλύτερα στοιχεία των δύο συστημάτων Mark III και NO96. Το βασικό στρώμα είναι κατασκευασμένο από INVAR πάχους 0,7cm και το δευτερεύον από TRIPLEX. Η πρωτοβάθμια και δευτεροβάθμια μόνωση αποτελείται από πάνελ αφρού Πολυουρεθάνης. Μέχρι στιγμής μόνο 3 πλοία με δεξαμενές τεχνολογίας CS1 έχουν κατασκευαστεί από ένα ναυπηγείο, αλλά διάφορα ναυπηγεία παγκοσμίως έχουν αποφασίσει να αναπτύξουν και να εξαπλώσουν την παραγωγή δεξαμενών τύπου MKIII & NO96.[45]

## Κεφάλαιο 2.2 Μέθοδοι μεταφοράς LNG

Στη συνέχεια θα αναφερθούμε εν συντομία στους τρόπους με τους οποίους το φυσικό αέριο μεταφέρεται από τον παραγωγό στον καταναλωτή. Το φυσικό αέριο καταλαμβάνει γενικά έναν μεγάλο όγκο αποθήκευσης σε θερμοκρασίες περιβάλλοντος [55], επομένως για αυτό τον λόγο είναι ενδεδειγμένη η μεταφορά του σε αέρια αμέσως αφότου εξάγεται από τον υπόγειο ταμειευτήρα. Υπάρχουν διάφορες μέθοδοι που χρησιμοποιούνται για να εξαγάγουν το φυσικό αέριο από πλατφόρμες εξόρυξης προς τις χώρες που το εισάγουν. Σύμφωνα με τους Thomas και Dawe (2003) αυτές οι μέθοδοι περιλαμβάνουν:

**(1)** Τη μεταφορά από πλοία μεταφοράς LNG, όπου το αέριο ψύχεται αρχικά σε μια υγρή κατάσταση σε περίπου  $-162^{\circ}\text{C}$  (το ατμοσφαιρικό σημείο βρασμού του) σε ατμοσφαιρική πίεση, και αντλείται έπειτα ως ρευστό στα καλά μονωμένα συστήματα συγκράτησης μέσα στο σκάφος ώστε να γίνει η μεταφορά. Για να χαμηλώσουν τη θερμοκρασία του μεθανίου σε αυτά τα επίπεδα γενικά υπάρχουν τρεις βασικές διαδικασίες ρευστοποίησης σήμερα και τις περιγράφουμε εν' συντομία εδώ:

- **Η Καθαρή διαδικασία ψυκτικών ουσιών** (Pure process) αυτή είναι παρόμοια σε γενικές γραμμές με τον κύκλο υγροποίησης αλλά προκειμένου να επιτευχθεί η χαμηλή θερμοκρασία που απαιτείται, περιλαμβάνονται τρία στάδια όπου το καθένα αποτελείται από την ψυκτική ουσία, το συμπιεστή και τους εναλλάκτες θερμότητας του. Το πρώτο στάδιο χρησιμοποιεί ψυκτική ουσία το προπάνιο, το δεύτερο είναι ένα στάδιο συμπύκνωσης που χρησιμοποιεί το αιθάνιο και τελικά το τρίτο στάδιο χρησιμοποιεί μεθάνιο. Η διαδικασία αυτή χρησιμοποιείται σε εγκαταστάσεις πριν από το 1970 με ένα βασικό πλεονέκτημα αυτής της μεθόδου να είναι η σχετικά μεγάλη εξοικονόμηση ενέργειας. Ένα μειονέκτημα είναι η μεγάλες σε έκταση εγκαταστάσεις που είναι απαραίτητες για τη διεξαγωγή της διαδικασίας, όπως και οι μεγάλες ποσότητες καθαρού αιθανίου και προπανίου που χρειάζονται.

- **Η Μικτή διαδικασία ψυκτικών ουσιών** (mixed process) ενώ με την καθαρή διαδικασία ψυκτικών ουσιών (όπως περιγράφεται ανωτέρω) έχουμε μια διαδικασία τριών κύκλων, με τη μικτή διαδικασία ψυκτικών ουσιών (συνήθως μεθάνιο, αιθάνιο, προπάνιο και άζωτο), η διαδικασία επιτυγχάνεται σε έναν κύκλο. Ο εξοπλισμός είναι λιγότερο σύνθετος από την καθαρή διαδικασία ψυκτικών ουσιών αλλά η κατανάλωση ισχύος είναι ουσιαστικά μεγαλύτερη και για αυτόν τον λόγο η χρήση της δεν είναι διαδεδομένη

- Η **μικτή διαδικασία ψυκτικών ουσιών με ταυτόχρονη ψύξη** (cooled mixed process) είναι η διαδικασία που είναι γενικά γνωστή ως διαδικασία MultiComponent και είναι ένας συνδυασμός των πιο πάνω διαδικασιών. Είναι η πιο κοινή διαδικασία που χρησιμοποιείται σήμερα.

**(2)** Μέσω συστημάτων σωληνώσεων για να μεταφερθεί το φυσικό αέριο σε αέρια κατάσταση δια μέσου ξηράς ή στο βυθό διά σε πίεση που κυμαίνεται μεταξύ 4,8 και 7,5 MPa. Αυτή η μέθοδος χρησιμοποιείται εκτενώς σε όλη την Ευρώπη, τις ΗΠΑ, τη Νότια Αμερική και τη Μέση Ανατολή

**(3)** Τη μεταφορά του υπό μορφή συμπιεσμένου φυσικού αερίου Compressed Natural Gas όπου το αέριο τοποθετείται σε ένα σκάφος με δεξαμενές πίεσης όπου βρίσκεται υπό υψηλή πίεση (25 MPa) πριν την μεταφορά του στα σημεία ζήτησης.

**(4)** Τη μεταφορά του υπό μορφή ένυδρης μορφής φυσικού αερίου (Natural Gas Hydrate -NGH). Το NGH είναι το προϊόν της μίξης του φυσικού αερίου με νερό για να διαμορφώσει έναν σταθερό κρυστάλλινο πάγο που μπορεί να μεταφερθεί από πλοία μεταφοράς χύδην φορτίου σε μεγάλες δεξαμενές τύπου φιαλών (thermos flask) και να αποθηκευτεί σε σχεδόν αδιαβατικές συνθήκες.

Από τα παραπάνω είναι προφανές πως η μεταφορά των υγραερίων γίνεται είτε σε πολύ χαμηλές θερμοκρασίες, που για ορισμένα φθάνει τους  $-157\text{ }^{\circ}\text{C}$ , είτε σε υψηλές πιέσεις. Αυτές οι συνθήκες παρουσιάζουν πολλές δυσχέρειες καθώς και ιδιαίτερους κινδύνους. Έτσι εξ αντικειμένου τα πλοία αυτά θεωρούνται πολύ επικίνδυνα σε ατυχήματα και για το λόγο αυτό τα μέτρα ασφαλείας που λαμβάνονται σε αυτά είναι πολύ σχολαστικά όπως προαναφέραμε. Η προσέγγιση τέτοιων πλοίων επιτρέπεται μόνο σε ειδικές προβλήτες συναφών εγκαταστάσεων που παρέχουν υψηλό βαθμό σε μέσα ασφαλείας, πρόβλεψης και αντιμετώπισης έκτακτων συνθηκών.

## Κεφάλαιο 2.3 Στατιστικά στοιχεία σχετικά με τα πλοία μεταφοράς LNG & προβλήματα σχεδίασης

Οι κύριες χρήσεις του LNG είναι να παραχθεί ηλεκτρική ενέργεια και να παραχθεί θερμογόνος ισχύς. Ένα από τα κύρια πλεονεκτήματα του φυσικού αερίου ως καύσιμο είναι ότι παράγει 50% λιγότερες εκπομπές CO<sub>2</sub> έναντι άλλων συμβατικών καυσίμων (Shin and Lee, 2009; ENGVA, 2006; Thomas and Dawe, 2003). Το μεθάνιο, ως κυρίαρχο συστατικό του φυσικού αερίου αποτελείται από ένα άτομο άνθρακα με τέσσερα άτομα του υδρογόνου (CH<sub>4</sub>), το οποίο είναι το απλούστερο μόριο υδρογονανθράκων. Ως εκ τούτου, η παραγωγή του διοξειδίου του άνθρακα θα μειωθεί πολύ έναντι της καύσης άλλων υδρογονανθράκων (π.χ. προπάνιο, βουτάνιο, κ.λ.π.) που έχουν περισσότερα άτομα άνθρακα. Επιπλέον, με τη νέα τεχνολογία, ο συνδυασμένος κύκλος αεριοστροβίλων με χρήση φυσικού αερίου ως καύσιμο έχει υψηλότερη απόδοση σε σχέση με την χρήση ατμοστροβίλων, για την παραγωγή της ηλεκτρικής ενέργειας με πιο μικρή κατανάλωση καυσίμων (MER, 2008c). Ένα ακόμα πλεονέκτημα είναι η αστάθεια στην τιμή του ακατέργαστου πετρελαίου όπου για παράδειγμα μόνο το 2008 αυξήθηκε σε περισσότερα από 100 δολάρια ανά βαρέλι (Miller, 2009a), καθιστώντας κατά συνέπεια το φυσικό αέριο όλο και περισσότερο ελκυστικό ως πηγή ενέργειας.

Η αύξηση σε ζήτηση για το φυσικό αέριο έχει ενισχύσει σημαντικά την αγορά LNG (Yamawaki, 2002 Batcheler, 2000). Πρόσφατες εκτιμήσεις της παγκόσμιας κατανάλωσης LNG αναδεικνύει μια αυξητική πορεία, με χαρακτηριστικό παράδειγμα μια αύξηση της τάξης του 29% το 2009 και το 2010. Η Μέση Ανατολή έχει τον κυρίαρχο ρόλο στην παραγωγή, που παρέχει 61% της συνολικής ποσότητας LNG που παράγεται ετησίως (Miller, 2009b). Η συνολική παγκόσμια παραγωγή έχει προβλεφθεί για να φθάσει τα 500 εκατομμύρια τόνους μέχρι το 2020 (Marzouqi, 2008). Η αύξηση σε ζήτηση λοιπόν οδηγεί εκεί που είναι ουσιαστικά αναπόφευκτη αύξηση στους αριθμούς της συνολικής χωρητικότητας των πλοίων μεταφοράς LNG.

Πριν από τη δεκαετία του '80 υπήρχαν μόνο 42 πλοία μεταφοράς LNG εν ενεργεία, μέχρι το 2002 ο αριθμός αυτός είχε αυξηθεί σε 129 (Kuver et Al, 2002) ενώ το 2008 υπήρχαν 291 πλοία μεταφοράς LNG εν ενεργεία και 98 σκάφη υπό κατασκευή (LWS, 2008). Εκτός από την αύξηση στα πλοία μεταφοράς, η μεμονωμένη μεταφορική ικανότητα φορτίου όγκου των νέων σκαφών έχει αυξηθεί επίσης γρήγορα. Αρχίζοντας από 150 m<sup>3</sup> το 1962, που αυξάνονται σε 27.400 m<sup>3</sup> το 1964 (MAN, 2009), σε 130.000 m<sup>3</sup> στις αρχές της δεκαετίας του '80 και σε 138.000 m<sup>3</sup> από τα μέσα της δεκαετίας του '90 (SB, 2009). Μέχρι το τέλος του 2005, το μέγιστο μέγεθος αυξήθηκε πάλι φτάνοντας τα 153.000 m<sup>3</sup> και συνέχισε μέχρι και 266.000 m<sup>3</sup> το 2009 (Motorship, 2009).

### Προβλήματα σχεδιασμού

Τα LNG carriers αποτελούνται από δύο βασικά μέρη, την γάστρα σε συνδυασμό με την εγκατάσταση πρόωσης, και το κρυογόνο τμήμα που διασφαλίζει την σωστή φύλαξη του φορτίου. Όλα τα LNG Carriers είναι απαραίτητο να σχεδιάζονται με διπλά τοιχώματα τόσο στον πυθμένα όσο και στα πλαϊνά του πλοίου ώστε σε περίπτωση σύγκρουσης ή προσάραξης, η ζημιά στις δεξαμενές φορτίου να είναι όσο το δυνατόν μικρότερη και να μην υπάρξει διαρροή (CSR & SOLAS). Επειδή το υγροποιημένο φυσικό αέριο έχει πολύ μικρή πυκνότητα (0,4 - 0,5 tn/m<sup>3</sup>), ένα πλοίο που σχεδιάζεται για την μεταφορά του θα έχει τεράστια χωρητικότητα σε σχέση με το DWT του. Σε αντίθεση με τα δεξαμενόπλοια πετρελαίου τα LNG Carriers σε κατάσταση ερματισμού δεν μπορούν να γεμίσουν τις δεξαμενές φορτίου με νερό. Μάλιστα πολλές φορές γίνεται κατακράτηση μιας μικρής ποσότητας LNG μέσα στις δεξαμενές (κατά την διάρκεια μιας κατάστασης ερματισμού) προκειμένου αυτές να διατηρήσουν την κανονική θερμοκρασία μεταφοράς.



## Κεφάλαιο 2.4 Κίνδυνοι και ατυχήματα σε LNG Tankers

### Κεφάλαιο 2.4.1 Κίνδυνοι

Οι κύριοι κίνδυνοι από LNG προκύπτουν από την κρυογόνο θερμοκρασία του, την ευφλεκτότητα του και τα χαρακτηριστικά διασποράς του ατμού που δημιουργείται. Σε υγρή μορφή, το LNG από μόνο του ούτε θα καεί ούτε θα εκραγεί. Το μεθάνιο, το κύριο συστατικό του LNG, είναι άχρωμο και άγευστο και είναι ταξινομημένο ως απλό ασφυξιογόνο στην ανθρώπινη έκθεση[46]. Το LNG εξατμίζεται ταχύτατα όταν εκτίθεται σε περιοχές θερμότητες όπως αυτές του θαλασσινού νερού, δημιουργώντας 620 με 630 m<sup>3</sup> φυσικού αερίου για κάθε 1 m<sup>3</sup> υγροποιημένου φυσικού αερίου. Όταν διαρρέυσει πάνω στο νερό, το LNG θα δημιουργήσει αρχικά ένα σύννεφο ατμού (οι κρύοι ατμοί είναι πυκνότεροι από τον αέρα και μένουν πιο κοντά στο νερό ή στο έδαφος). Καθώς αναμιγνύεται το σύννεφο αυτό με τον αέρα, θερμαίνεται και διασκορπίζεται στην ατμόσφαιρα. Εάν δεν υπάρξει ανάφλεξη, το εύφλεκτο σύννεφο ατμού θα παρασυρθεί από τον αέρα έως ότου η διασπορά του θα αραιώσει τους ατμούς σε επίπεδα κάτω από την ελάχιστη συγκέντρωση που τους κάνει εύφλεκτους. Σε μια συγκέντρωση 5% του αερίου στον αέρα, οι ατμοί LNG βρίσκονται στο κατώτερο όριο ανάφλεξης τους (lower flammability limit). Κάτω από αυτή την αναλογία ατμού- αέρα, το σύννεφο είναι πολύ αραιωμένο για να αναφλεγεί. Σε ένα ποσοστό 15% αυτής της αναλογίας, οι ατμοί LNG βρίσκονται στο ανώτερο εύφλεκτο όριο τους και είναι η κατάσταση πέρα της οποίας το σύννεφο είναι πολύ πλούσιο, κορεσμένο και έτσι ανίκανο για ανάφλεξη.

Η απόσταση που μπορεί να φτάσει ένα εύφλεκτο σύννεφο ατμού LNG εξαιτίας του αέρα, είναι μια συνάρτηση του όγκου του υγρού που διέρρευσε, του ρυθμού της διαρροής αυτής και των καιρικών συνθηκών που επικρατούν εκείνη τη στιγμή. Επίσης, προκειμένου να απομακρυνθεί το σύννεφο από το σημείο της διαρροής, θα πρέπει να αποφύγει την ανάφλεξη. Διαπιστώνεται, ότι οι μεγάλες απελευθερώσεις φορτίου από ένα LNG πλοίο θα απαιτούσαν πιθανώς μια σημαντική πηγή ενέργειας για ανάφλεξη (να παραβιαστεί δηλαδή το εξωτερικό τμήμα της γάστρας, το εσωτερικό μέρος του double hull και η δεξαμενή αποθήκευσης). Ένα γεγονός ικανοποιητικού μεγέθους ώστε να παραβιαστεί η δεξαμενή του LNG, μπορεί να παρέχει τις πηγές ανάφλεξης του ίδιου του φορτίου. Εάν το εύφλεκτο σύννεφο αναφλέγεται από το γεγονός της έναρξης του ρήγματος ή από κάποιες άλλες πηγές (π.χ. στο σκάφος, σε άλλα κοντινά σκάφη, στην ακτή), η φλόγα θα κατευθυνθεί προς την πηγή παραγωγής του ατμού και το εύφλεκτο σύννεφο δεν θα απομακρυνθεί πολύ από το σημείο εκείνο. Όταν ένα εύφλεκτο μίγμα αερίου-αέρα από μια διαρροή σε πλοίο μεταφοράς LNG αναφλέγεται, μπορεί να οδηγήσει σε μια πυρκαγιά λάμψης (flash fire), η οποία είναι μια πυρκαγιά μικρής διάρκειας που καίει

τα αέρια που ήδη αναμίχθηκαν με τον αέρα και υπάρχουν σε αυτόν σε εύφλεκτες συγκεντρώσεις. Το μέτωπο της φλόγας κατευθύνεται προς τα πίσω στη περιοχή της διαρροής, υπό την προϋπόθεση ότι η συγκέντρωση αερίου κατά μήκος αυτής της πορείας είναι αρκετά υψηλή να συνεχίσει. Ο ρυθμός με τον οποίο το μέτωπο αυτό ταξιδεύει εξαρτώμενο από το άκαυστο αέριο ονομάζεται στρωτή ταχύτητα ανάφλεξης (laminar burning velocity).

Ένα μίγμα μεθανίου- αέρα θα καίγεται αργά, τείνοντας να αναφλέξει τα καύσιμα συστατικά του σύννεφου του αερίου, ενώ οι γρήγορες ταχύτητες μετάδοσης της φλόγας οδηγούνται περισσότερο σε πυρκαγιές λάμψης. Αν και οι ατμοί LNG μπορούν να εκραγούν (δημιουργώντας τεράστιες υπερπιέσεις), εάν υπάρξει ανάφλεξη μέσα σε περιορισμένο χώρο, όπως ένα κτίριο ή μια κατασκευή, δεν υπάρχει καμία ένδειξη που να προτείνει ότι το LNG εκρήγνυνται σε περίπτωση ανάφλεξης σε ανοιχτό χώρο. Έχουν γίνει πειράματα σχετικά με τον καθορισμό του πότε μπορεί να εκραγούν τα μίγματα μεθανίου- αέρα και μέχρι σήμερα είναι αρνητικά.

Οι κύριοι κίνδυνοι από το LNG σχετίζονται με την εύφλεκτη διασπορά του ατμού και την θερμική ακτινοβολία που προκαλείται. Δευτερογενείς κίνδυνοι, όπως κρουγόνα εγκαύματα και ασφυξία, είναι χαρακτηριστικά που εντοπίζονται στις διαδικασίες μεταφοράς και κατά την αποθήκευση του LNG. Το LNG είναι λιγότερο επικίνδυνο από το υγροποιημένο αέριο πετρελαίου (LPG) και το υγροποιημένο αιθυλένιο, τα οποία έχουν μεγαλύτερο ειδικό βάρος, μεγαλύτερη τάση στο να δημιουργήσουν εκρηκτικά σύννεφα ατμού και απαιτούν λιγότερη ενέργεια για την ανάφλεξη τους. Το LNG δεν είναι τοξικό και εξατμίζεται ακαριαία, επομένως οι μακροπρόθεσμες περιβαλλοντικές επιδράσεις από μια απελευθέρωση είναι αμελητέες εάν δεν υπάρχει καμία ανάφλεξη του νέφους αερίου που δημιουργείται.

Παρακάτω παρουσιάζονται οι διάφοροι κίνδυνοι που σχετίζονται με μια διαρροή υγροποιημένου φυσικού αερίου στο νερό, όπως τους έχει κατηγοριοποιήσει ο ABS (2004).

- Κίνδυνοι πυρκαγιάς
- Εκρήξεις
- Άμεση μεταβατική φάση (RPT)
- Κρουγόνες επιδράσεις

## A. ΚΙΝΔΥΝΟΙ ΠΥΡΚΑΓΙΑΣ

Το LNG ατμοποιείται γρήγορα δεδομένου ότι απορροφά θερμότητα από το περιβάλλον, και το νέφος ατμών που προκύπτει είναι εύφλεκτο όταν αναμειγνύεται στον αέρα στις συγκεντρώσεις από 5 σε 15% (κατά όγκο) όπως προαναφέραμε. Οι ιδιότητες του σχετικά με τη φωτιά είναι συγκρίσιμες με άλλα ελαφριά καύσιμα υδρογονανθράκων. Η μόνη σημαντική διαφορά είναι ότι το μοριακό βάρος του είναι αρκετά μικρότερο από αυτό του αέρα, έτσι μόλις θερμανθεί πάνω από περίπου  $-162\text{ }^{\circ}\text{C}$  θα γίνει λιγότερο πυκνό από τον αέρα και θα τείνει να αυξηθεί και να διασκορπίσει γρηγορότερα. Πιο συγκεκριμένα, το αέριο συμπύκνωμα του LNG στη θερμοκρασία περιβάλλοντος είναι ελαφρύτερο από τον αέρα και συγκεκριμένα το ειδικό βάρος του σε σχέση με τον αέρα είναι 0,55 μόλις επιτευχθεί θερμική ισορροπία (ABS Consulting, 2004).

Υπάρχουν 3 είδη πυρκαγιών που μπορεί να προκύψουν από το φυσικό αέριο: Η πυρκαγιά λίμνης (Pool Fire), η πυρκαγιά πίδακα (Jet Fires) και οι πυρκαγιές λάμψης (Flash Fires). Τα βασικά χαρακτηριστικά των 2 πρώτων παρουσιάζονται εν συντομία παρακάτω:

### Pool Fire

Όταν ένα εύφλεκτο υγρό απελευθερώνεται από μια δεξαμενή αποθήκευσης ή μια σωλήνωση, μπορεί να δημιουργηθεί μια υγρή εύφλεκτη λίμνη. Καθώς σχηματίζεται αυτή η λίμνη, ένα μέρος του υγρού θα εξατμιστεί, και αν το εύφλεκτο αυτό σύννεφο βρει μια πηγή ανάφλεξης, η φλόγα μπορεί να ταξιδέψει προς τα πίσω στη διαρροή, προκαλώντας μια πυρκαγιά στη λίμνη, πάνω από την οποία καίγεται το σύννεφο του αερίου που μόλις είχε εξατμιστεί από αυτήν και αναμίχθηκε με τον αέρα.

### Jet fire

Εάν τα συμπιεσμένα ή υγροποιημένα αέρια διαρρεύσουν από την δεξαμενή αποθήκευσης ή από διάφορες σωληνώσεις υπό υψηλή πίεση τα υλικά που απελευθερώνονται από την τρύπα θα σχηματίσουν ένα gas jet (πίδακα αερίου) που παρασύρεται και αναμειγνύεται με τον αέρα που το περιβάλλει. Εάν το υλικό αυτό συναντήσει κάποια πηγή ανάφλεξης, μπορεί να δημιουργηθεί μια πυρκαγιά τύπου πίδακα. Για το LNG που αποθηκεύεται σε χαμηλή πίεση σαν υγρό, όπως γίνεται στα Liquefied Natural Gas Carriers, τέτοιου είδους πυρκαγιά είναι απίθανη. Μπορούν μόνο να συμβούν κατά τη διάρκεια της εκφόρτωσης ή της μεταφοράς όταν οι πιέσεις είναι αυξημένες λόγω της άντλησης.

## **B.ΕΚΡΗΞΕΙΣ**

Όπως αναφέρθηκε και παραπάνω, μια πυρκαγιά λάμπσης μπορεί να εμφανιστεί εάν απελευθερωθεί LNG στην ατμόσφαιρα και αναφλεγεί. Εάν συμβεί αυτό σε ανοιχτές περιοχές, το μεθάνιο δεν είναι γνωστό ότι θα δημιουργήσει καταστρεπτικές υπερπιέσεις. Παρόλα αυτά, εάν υπάρχει κάποιος περιορισμός στο σύννεφο του αερίου, το μεθάνιο μπορεί να προκαλέσει τέτοιες καταστρεπτικές υπερπιέσεις που τελικά θα οδηγήσουν σε έκρηξη. Αυτός ο περιορισμός μπορεί να δημιουργηθεί από κενά διαστήματα μέσα στο πλοίο ή σε γειτονικές κατασκευές, όπως κτίρια στη ξηρά ή διπλανά πλοία. Περιοχές ασφυκτικά γεμάτες με εξοπλισμό και κατασκευές μπορούν επίσης να διευκολύνουν τη δημιουργία καταστρεπτικών υπερπιέσεων εάν ένα σύννεφο αερίου αναφλεγεί μέσα σε αυτές.

Παραδείγματος χάριν εάν ένα σύννεφο ατμού διεισδύσει σε εγκαταστάσεις χημικών διαδικασιών σε μια περιοχή με διάφορα σκάφη, κατασκευές, και σωληνώσεις και το σύννεφο αναφλεγεί, το τμήμα του σύννεφου μέσα σε εκείνη την κορεσμένη περιοχή μπορεί να παραγάγει αυτές τις υπερπιέσεις. Ένα μεγαλύτερο μέρος όγκου των βαρύτερων υδρογονανθράκων του LNG

**(1)** μειώνει την ελάχιστη ενέργεια ανάφλεξης που απαιτείται για την εκपुरσοκρότηση και

**(2)** αυξάνει το ειδικό βάρος του μίγματος υδρογονανθράκων (και ως εκ τούτου μειώνει την τάση να διασκορπιστεί γρήγορα).

Και τα δύο αυτά αποτελέσματα αυξάνουν την πιθανότητα της εμφάνισης υπερπιέσεων [47]. Απαραίτητη προϋπόθεση για να συμβεί κάτι τέτοιο, είναι η διαρροή του υγρού να γίνει σε ένα περιορισμένο χώρο με την ύπαρξη εμποδίων. Μονάχα σε εκείνη την περίπτωση εμφανίζεται το φαινόμενο της έκρηξης, κάτι το οποίο είναι εξαιρετικά δύσκολο να συμβεί, παρόλα αυτά όμως πρέπει να εξεταστεί και το ενδεχόμενο αυτό όσο μικρό και αν είναι. Μέχρι στιγμής δεν έχει παρατηρηθεί τέτοιο συμβάν σε υγραεριοφόρα καθώς κατά ένα συντριπτικό ποσοστό το υγρό διαρρέει στη θάλασσα και όχι σε εσωτερικά τμήματα του πλοίου.

## **Γ.ΑΜΕΣΗ ΜΕΤΑΒΑΣΗ ΦΑΣΗΣ (RAPID PHASE TRANSITION)**

Η άμεση μετάβαση φάσης είναι ο όρος που χρησιμοποιείται για να περιγράψει ένα φαινόμενο που αναγνωρίζεται σε μερικά πειράματα απελευθέρωσης LNG, που περιλαμβάνουν τη σχεδόν στιγμιαία μετάβαση του υγρού στη φάση ατμού και μια σχετικά γρήγορη αύξηση πίεσης. Όταν το LNG διαμορφώνει μια λίμνη στο νερό, η θερμότητα από το νερό το ατμοποιεί γρήγορα, εντούτοις, αυτός ο βρασμός δεν είναι το φαινόμενο καλούμενο RPT. Σε μια κατάσταση RPT, ένα μέρος του LNG αλλάζει από την υγρή στην αέρια μορφή ουσιαστικά στιγμιαία. Αν και ο φυσικός μηχανισμός δεν έχει γίνει απόλυτα κατανοητός, η RPT αποδίδεται στην

υπερθέρμανση του LNG. Μια κατάσταση RPT μπορεί να οδηγήσει σε δύο τύπους αποτελεσμάτων:

- (1) Υπερπίεση ως αποτέλεσμα της γρήγορης αλλαγής φάσης και
- (2) Διασπορά της «ριπής» του LNG που αποβάλλεται στην ατμόσφαιρα.

Οι γρήγορες αλλαγές φάσης έχουν προκαλέσει πολυάριθμες εκρήξεις αερίου σε χυτήρια και σε άλλες βιομηχανικές διαδικασίες, αλλά δεν έχουν οδηγήσει σε σημαντικά ατυχήματα σχετικά με τη μεταφορά LNG [47]. Τα γεγονότα άμεσης μετάβασης φάσης μπορεί να είναι αρκετά βίαια και να εκτινάξουν τμήμα του υγρού στον αέρα σε μορφή ριπής. Ένα μεγάλο μέρος του LNG θα εκτιναχθεί και στη συνέχεια θα εξατμιστεί καθώς θα πέφτει προς τα κάτω. Η πηγή από την οποία παράγεται το αέριο θα είναι διαφορετική συγκριτικά με μια ομαλά διαδιδόμενη λίμνη, με πιθανές μεγάλες επιδράσεις στην διασπορά του σύννεφου στον αέρα. Σε αυτές τις περιπτώσεις, η επικύρωση μοντέλων θεωρητικών και υπολογιστικών σχετικά με ομαλή εξάπλωση και ατμοποίηση της λίμνης σε πειράματα μεγάλης κλίμακας, μπορεί να μην είναι εφικτή. Η RPT δημιουργεί ένα εκρηκτικό κύμα, το οποίο μπορεί να προκαλέσει ζημιές στις παρακείμενες περιοχές. Παρά όλα αυτά όμως, όλες οι ενδείξεις οδηγούν στο συμπέρασμα ότι δεν θα υπάρξει κάποιο ιδιαίτερα σοβαρό πρόβλημα σε περίπτωση που υπάρξει RPT σε πλοία LNG με γάστρα από χάλυβα

#### **Δ. ΚΡΥΟΓΟΝΕΣ ΕΠΙΔΡΑΣΕΙΣ**

Τα LNG υγραεριοφόρα σχεδιάζονται ώστε να αποτραπεί η επαφή των εσωτερικών και εξωτερικών τοιχωμάτων της γάστρας με το LNG. Είναι δυνατό με μια απελευθέρωση υγρού LNG στο εσωτερικό τοίχωμα να προκαλούσε χαμηλές θερμοκρασίες σε περιοχές του τοιχώματος που δε σχεδιάζονται για αντοχή σε τόσο χαμηλές θερμοκρασίες, δημιουργώντας διάφορα προβλήματα όπως αυτό του κινδύνου Ψαθυρής Θραύσης.

Μια μελέτη από τον Lloyds (2001) περιλαμβάνει συνοπτικές περιγραφές 10 περιπτώσεων διαρροών LNG από LNGC που εμφανίστηκαν μεταξύ 1965 και 1989. Επτά από αυτές τις δέκα οδήγησαν στο σπάσιμο των καλυμμάτων του καταστρώματος ή της δεξαμενής αποθήκευσης, αλλά κανένα από αυτά δεν προκάλεσε τη σοβαρή δομική ζημία. Λαμβάνοντας υπόψη τα σημεία όπου εμφανίστηκε η ζημία (δηλ. κάλυμμα δεξαμενής ή καταστρώματος), είναι πιθανό ότι όλες αυτές οι απελευθερώσεις να εμφανίστηκαν από τα συστήματα σωληνώσεων που χρησιμοποιήθηκαν κατά τη διάρκεια της μεταφοράς του LNG στις δεξαμενές. Επίσης, τα σκάφη αυτά που κατασκευάζονται από το 1976 πρέπει να συμμορφώνονται με τους εκάστοτε κατασκευαστικούς κανόνες του διεθνούς

οργανισμού IMO και να χρησιμοποιείτε χάλυβας υψηλής αντοχής που απαιτείται για τις χαμηλές αυτές θερμοκρασίες.

Κάποιος μπορεί να ανατρέξει στους κανονισμούς ασφαλείας για τα πλοία μεταφοράς LNG που αφορούν την κατασκευή, τον εξοπλισμό και τις λειτουργίες τους όπως αυτοί περιγράφονται στους "Gas Codes" του IMO. Οι τρεις "κώδικες αερίου" που έχουν δημιουργηθεί για την μεταφορά LNG είναι [55]

- 1.** Ο κώδικας για τα υπάρχοντα πλοία που μεταφέρουν υγροποιημένο αέριο χύδην Code for existing ships carrying liquefied gases in bulk (1976 Edition).
- 2.** Ο κώδικας κατασκευής και εξοπλισμού των πλοίων που μεταφέρουν υγροποιημένο αέριο χύδην GC Code (1983 Edition).
- 3.** Ο διεθνής κώδικας κατασκευής και εξοπλισμού των πλοίων που μεταφέρουν υγροποιημένο αέριο χύδην IGC Code (2016 Edition).

## Κεφάλαιο 2.4.2 Αξιολόγηση ρίσκου σε πλοία μεταφοράς LNG

Η ασφάλεια και η αξιοπιστία των υγραεριοφόρων LNG μέχρι στιγμής ήταν εξαιρετικές στον κλάδο της ναυτιλίας, καθώς έχει συμβεί μικρός αριθμός ατυχημάτων με περιορισμένες καταστροφές κάτι το οποίο επιτεύχθηκε ως αποτέλεσμα των χαρακτηριστικών ασφαλείας που ισχύουν για την αποφυγή αδικαιολόγητης απελευθέρωσης LNG από τους χώρους φορτίου αλλά και από άλλες συναφείς θαλάσσιες εγκαταστάσεις όπως οι δεξαμενές .

Η εκτίμηση κινδύνου είναι μια συνολική εκτίμηση της πιθανότητας και του βαθμού των πιθανών συνεπειών σε μια επικίνδυνη κατάσταση προκειμένου να επιλεγούν τα κατάλληλα μέτρα ασφαλείας (Wang and Trbojevic, 2007). Πριν από την αξιολόγηση κινδύνου, όλα τα εμπλεκόμενα μέρη πρέπει να έχουν κοινή αντίληψη για τους στόχους της εκτίμησης τις μεθόδους που θα χρησιμοποιηθούν, τους απαιτούμενους πόρους και τον τρόπο με τον οποίο θα εφαρμοστούν τα αποτελέσματα (ABS, 2000).

Η εκτίμηση επικινδυνότητας των LNG Carriers μπορεί να πραγματοποιηθεί χρησιμοποιώντας ποιοτική ή ποσοτική ανάλυση κινδύνου, ανάλογα με τις απαιτήσεις των αναλυτών ασφαλείας LNG και τα διαθέσιμα ιστορικά δεδομένα για τα περιστατικά ατυχημάτων που αφορούν LNG Carriers. Η διαδικασία μπορεί να είναι προληπτική και περιλαμβάνει νέους κινδύνους που εκτιμώνται βάσει της βελτίωσης της τεχνολογίας των πλοίων μεταφοράς LNG. Από τις προηγούμενες μελέτες εκτίμησης κινδύνου σε LNG Carriers έχει διαπιστωθεί ότι οι κίνδυνοι που σχετίζονται με τις εργασίες που εκτελούνται από LNG Carriers διαχειρίζονται / μειώνονται σε μεγάλο βαθμό χρησιμοποιώντας χαρακτηριστικά ασφαλείας όπως η κατασκευή με χρήση διπλής γάστρας, χρήση χάλυβα υψηλής αντοχής σε κρίσιμες περιοχές, σύστημα σωληνώσεων διπλού τοιχώματος, συστήματα υψηλής ευαισθησίας ανίχνευσης διαρροών αερίων, καλά καταρτισμένο και εξειδικευμένο πλήρωμα, αδρανείς διασταυρωμένοι χώροι, συστήματα ψεκασμού νερού, ξηρά χημικά, συστήματα πυρόσβεσης CO<sub>2</sub> και συστήματα έκτακτης εκφόρτωσης φορτίου που ενσωματώνονται στα LNGC. Ανεξάρτητα από τα ήδη υπάρχοντα χαρακτηριστικά ασφαλείας για τα LNGC και τις άλλες εγκαταστάσεις LNG που συνδέονται με αυτά, οι κίνδυνοι που ενδέχεται να επηρεάσουν την καλή λειτουργία των συστημάτων μεταφοράς και των υποσυστημάτων του LNG μπορούν να συρρικνωθούν, ιδίως με την ανάπτυξη και την εφαρμογή νέων τεχνολογιών. Οι ειδικοί της βιομηχανίας υδροποιημένου φυσικού αερίου μπορούν να διεξάγουν πλήρη εκτίμηση κινδύνων για τα LNGC και άλλες συναφείς θαλάσσιες εγκαταστάσεις χρησιμοποιώντας τεχνικές ανάλυσης ασφαλείας / κινδύνου όπως:

Στατιστική Ανάλυση Ναυτικών Ατυχημάτων σε πλοία LNG-LPG  
και Ανάπτυξη Αλγορίθμου Εξόρυξης Δεδομένων

- Event Tree Analysis (ETA).
- Risk Matrix.
- Fault Tree Analysis (FTA).
- Failure Mode, Effect and Critical Analysis (FMECA).
- Preliminary Hazard Analysis (PHA).
- Hazard and Operability study (HAZOP).
- What If Technique.
- Cause Consequence Analysis (CCA).



## ΚΕΦΑΛΑΙΟ 3. Αξιολόγηση και έλεγχος των δεδομένων με χρήση της μεθόδου $\chi^2$

### Κεφάλαιο 3.1 Απαραμετρική Συμπεριστατολογία

Πολλά από τα κριτήρια της στατιστικής συμπερασματολογίας βασίζονται σε περιοριστικές υποθέσεις για την μορφή της κατανομής του πληθυσμού από τον οποίο λαμβάνονται οι παρατηρήσεις. Οι διάφορες παράμετροι της κατανομής όπως η μέση τιμή και η διασπορά παραμένουν άγνωστες και συνήθως ενδιαφερόμαστε να εξάγουμε διάφορα συμπεράσματα για τις παραμέτρους από την πληροφορία που λαμβάνουμε από ένα τυχαίο δείγμα παρατηρήσεων.[49] Για παράδειγμα, βάσει ενός τυχαίου δείγματος από την υποτιθέμενη κατανομή του χαρακτηριστικού που εξετάζουμε ενδιαφερόμαστε για εκτίμηση ορισμένων παραμέτρων σε σημείο ή σε διαστήματα εμπιστοσύνης, είτε για κάποιο έλεγχο υπόθεσης για τις διάφορες παραμέτρους. Παραδείγματα τέτοιων υποθέσεων είναι:

1. Μια κανονική κατανομή έχει συγκεκριμένο μέσο όρο ή/και διασπορά.
2. Δύο κανονικές κατανομές έχουν ίσες μέσες τιμές ή/και διασπορές.
3. Κατά μέσο όρο ο αριθμός των ασθενών που καταφθάνουν σ' ένα νοσοκομείο που εφημερεύει δεν ξεπερνά τους δέκα ανά ώρα.

Από τις περιοριστικές υποθέσεις που τίθενται για τη μορφή της κατανομής του εξεταζόμενου χαρακτηριστικού, συνηθέστερη είναι η υπόθεση της κανονικότητας. Επιπλέον είναι δυνατόν να υποθέσουμε ότι οι πληθυσμοί που εξετάζουμε έχουν ίσες διασπορές ή ότι υπάρχει μια γνωστή σχέση μεταξύ των διασπορών [49]. Στην περίπτωση της κανονικότητας το κριτήριο που χρησιμοποιείται συνήθως προκύπτει σαν συνάρτηση της μέσης τιμής και της διασποράς καθόσον μια κανονική κατανομή χαρακτηρίζεται πλήρως από την μέση τιμή και τη διασπορά της. Διαφορετικές υποθέσεις για τη μορφή της κατανομής που συνήθως γίνονται, είναι της γάμμα κατανομής, της εκθετικής κατανομής, της ομοιόμορφης κατανομής, κ.λ.π. Από τις διακριτές κατανομές συνήθως χρησιμοποιούνται η διωνυμική και η κατανομή Poisson [50] (όπως π.χ. στην παραπάνω υπόθεση (3) όπου είναι δυνατό να υποθέσουμε ότι ο αριθμός των ασθενών που καταφθάνουν σε μια ώρα σε ένα νοσοκομείο ακολουθεί την κατανομή Poisson). Τέτοιες υποθέσεις που προσδιορίζουν τη μορφή της κατανομής των παρατηρήσεων ονομάζονται **παραμετρικές**. Τα αντίστοιχα κριτήρια που χρησιμοποιούνται για τον έλεγχο παραμέτρων της υποτιθέμενης κατανομής καλούνται **παραμετρικά** ή **κλασσικά** κριτήρια. Τέλος, το κομμάτι εκείνο της στατιστικής συμπερασματολογίας που διαπραγματεύεται τέτοια κριτήρια αποτελεί τη λεγόμενη **παραμετρική συμπερασματολογία**.

Σε πολλές όμως περιπτώσεις οι υποθέσεις αυτές για τη μορφή της κατανομής του γεννήτορος πληθυσμού είναι δύσκολο να αιτιολογηθούν. Σε άλλες επίσης περιπτώσεις η μορφή της κατανομής μπορεί να είναι γνωστή αλλά δεν είναι κατάλληλη (εύχρηστη) για τη χρησιμοποίηση των τυπικών μεθόδων της στατιστικής συμπερασματολογίας. Τέλος, σε άλλες περιπτώσεις τα ίδια τα δεδομένα εκ των πραγμάτων δεν ικανοποιούν (πληρούν) τις υποθέσεις. Είναι γνωστό ότι μερικές παραμετρικές διαδικασίες γίνονται αναξιόπιστες όταν εφαρμοστούν με σχετικά μικρές αποκλίσεις από τη μορφή της υποτιθέμενης κατανομής αν και είναι στιβαρές (robust) σε αποκλίσεις από την κανονικότητα. Όπως π.χ. στην περίπτωση κανονικών κατανομών όπου το κριτήριο F για τον έλεγχο διασπορών είναι πολύ ευαίσθητο σε αποκλίσεις της υποτιθέμενης κανονικής κατανομής.

Ένα μεγάλο πρόβλημα που εμφανίζεται είναι το κατά πόσο μπορούμε να εφαρμόσουμε στην πράξη παραμετρικά κριτήρια όταν η υποτιθέμενη μορφή της κατανομής δεν ευσταθεί. Εξαιτίας των πολλών θεωρητικών αποτελεσμάτων για την κανονική και της ευχρηστίας που παρατηρείται, υπάρχει η τάση να θεωρούμε πολλές κατανομές σαν κανονικές και να χρησιμοποιούμε τα αποτελέσματα της κανονικής θεωρίας. Σε τέτοιες περιπτώσεις συνήθως τίθεται το ερώτημα της ευστάθειας (robustness) μιας στατιστικής διαδικασίας, δηλαδή ο καθορισμός του μεγέθους της ισχύος των μεθόδων που πηγάζουν από την κανονική θεωρία.

Σε τέτοιες περιπτώσεις, όπως και σε περιπτώσεις που έχουμε κατηγορικά ή διατάξιμα δεδομένα (επιλογή κάποιου χρώματος, προτίμηση φαγητού, ταξινόμηση ασθενών αναφορικά με την αντίδρασή τους π.χ. σε μια απεργία του νοσηλευτικού προσωπικού) γίνεται αντιληπτό ότι απαιτούνται άλλες μέθοδοι στατιστικής συμπερασματολογίας που να μην βασίζονται σε λεπτομερείς υποθέσεις για την κατανομή του πληθυσμού από τον οποίο λαμβάνονται οι παρατηρήσεις. Συνήθως γίνονται πολύ ασθενείς υποθέσεις για την κατανομή των παρατηρήσεων (π.χ. υπόθεση συνέχειας ή συμμετρικότητας) χωρίς πάντως να προσδιορίζεται καμία συγκεκριμένη μορφή για την κατανομή.[48] Τέτοιες μέθοδοι καλούνται **ανεξάρτητες κατανομών (distribution-free)** καθόσον η ισχύς τους δεν εξαρτάται από τη μορφή της υποκείμενης κατανομής. Σκοπός μας δηλαδή στην περίπτωση αυτή είναι να βρούμε στατιστικές μεθόδους που να εφαρμόζονται σε μια ευρεία κατηγορία κατανομών όπως π.χ. σ' όλες τις συνεχείς κατανομές. Η στατιστική που ασχολείται με τέτοια προβλήματα καλείται και **απαραμετρική στατιστική**, τα δε αντίστοιχα κριτήρια **απαραμετρικά**. Οι έννοιες **"απαραμετρική"** (non-parametric) και **"ελεύθερη κατανομή"** (distribution-free) δεν είναι ακριβώς ταυτόσημες. Ο όρος "μη παραμετρική" αναφέρεται στην περιγραφή του προβλήματος, ενώ ο όρος "ελεύθερη κατανομή" αναφέρεται στη μέθοδο που χρησιμοποιείται για τη λύση του προβλήματος. Στη διεθνή βιβλιογραφία πάντως δεν υπάρχει γενικά δεκτός ορισμός των όρων αυτών οπότε χρησιμοποιούνται και οι δύο ισοδύναμα, χωρίς δηλαδή να διαχωρίζονται αυστηρά οι έννοιες τους.

Μερικά ενδεικτικά παραδείγματα μη παραμετρικών στατιστικών τεστ είναι τα εξής [48]:

**1.Διωνυμικό Κριτήριο (Binomial Test).** Η μηδενική υπόθεση αυτής της μεθόδου συνήθως είναι ότι τα ενδεχόμενα είναι ισοπίθανα και χρησιμοποιείται για έλεγχο δεδομένων τα οποία χωρίζονται σε δύο κατηγορίες.

**2.Κριτήριο Kolmogorov-Smirnov (K-S).** Η μηδενική υπόθεση αυτής της μεθόδου συνήθως είναι ότι η κατανομή των δεδομένων ακολουθεί την εμπειρική κατανομή (K-S) και χρησιμοποιείται για να ελέγξει συνεχείς μονοδιάστατες κατανομές πιθανοτήτων.

**3.Κριτήριο Crammer-Von Mises.** Η μηδενική υπόθεση αυτής της μεθόδου είναι ότι η συνάρτηση αθροιστικής κατανομής του δείγματος ακολουθεί την εμπειρική συνάρτηση κατανομής Crammer – Von Misses και χρησιμοποιείται για να ελέγξει το goodness of fit ενός συνόλου παρατηρήσεων.

**4.Κριτήριο  $\chi^2$**  Η μηδενική υπόθεση αυτής της μεθόδου είναι ότι τα δεδομένα είναι ανεξάρτητα. Μπορεί να χρησιμοποιηθεί για να καθοριστεί αν η κατανομή ενός συνόλου δεδομένων αποκλίνει από την αναμενόμενη αν αυτά είναι ανεξάρτητα.

Από αυτά τα 4 κριτήρια θα χρησιμοποιήσουμε το τελευταίο (4) για να προσδιορίσουμε κατά πόσο κάποια από τα δεδομένα τα οποία διαθέτουμε επηρεάζουν ένα κομμάτι του αποτελέσματος, τον τύπο ατυχήματος.

## Κεφάλαιο 3.2 Έλεγχος ανεξαρτησίας $\chi^2$ (Chi square Independence test)

Το κριτήριο  $\chi^2$  χρησιμοποιείται όταν θέλουμε να ελέγξουμε κατά πόσο μια εμπειρική κατανομή συχνοτήτων συμφωνεί με μια υποθετική κατανομή για διακριτά δεδομένα και αποτελέσματα. Η θεωρητική κατανομή μπορεί να είναι τελείως ορισμένη είτε μερικές παράμετροι να είναι άγνωστες που πρέπει να εκτιμηθούν από τα δεδομένα. Οι  $n$  παρατηρήσεις  $\chi_1, \chi_2, \dots, \chi_n$  ταξινομούνται σε ( $k \geq 2$ ) ξένες μεταξύ τους κατηγορίες (κελία) και έστω  $n_i$  ο αριθμός των παρατηρήσεων που ανήκουν στο κελί  $i, i=1,2,\dots,k$ . Προφανώς ισχύει:

$$n_1+n_2+\dots+n_k = n$$

Όταν η υποθετική κατανομή είναι γνωστή τότε προσδιορίζεται η πιθανότητα  $P_i$  μια παρατήρηση να "πέσει" στην κατηγορία  $i$ , όπου ισχύει:

$$P_1+P_2+\dots+P_k = 1$$

Αν καλέσουμε  $E$  μια τυχαία κατανομή των παρατηρήσεων στα  $k$ -κελία δηλαδή:

$$E=\{n_1, n_2, \dots, n_k\} \text{ τότε ισχύει: } P(E) = \frac{n!}{n_1!n_2!\dots n_k!} P_1^{n_1} P_2^{n_2} \dots P_k^{n_k}$$

δηλαδή έχουμε την πολυωνυμική κατανομή με παραμέτρους  $n$  και  $P_1, P_2, \dots, P_k$ .

Οι αναμενόμενες συχνότητες στα  $k$  κελία είναι τότε:  $E_i = nP_i, i = 1, 2, \dots, k$

Το πρόβλημα της προσαρμογής μιας υποτιθέμενης κατανομής μετατρέπεται σε έλεγχο υπόθεσης για τις πιθανότητες  $P_1, P_2, \dots, P_k$  πολυωνυμικής κατανομής. Έχουμε δηλαδή να ελέγξουμε την υπόθεση:

$H_0 : P_1=P_1, P_2 = P_2, \dots, P_k = P_k$  όπου  $P_1, P_2, \dots, P_k$  είναι γνωστοί αριθμοί με

$$P_i \geq 0, i = 1,2,\dots,k \text{ και } P_1 + P_2 + \dots + P_k = 1$$

Η εναλλακτική υπόθεση της  $H_0$  είναι :

$H_1 : \text{ισχύει } P_i \neq P_i \text{ για ένα τουλάχιστον } i, i = 1, 2, \dots, k.$

Ισοδύναμα, ο έλεγχος της  $H_0$  κατά της  $H_1$ , μετατρέπεται σε έλεγχο του κατά πόσο οι παρατηρούμενες συχνότητες  $n_i = 0_i$  διαφέρουν από τις αναμενόμενες συχνότητες

$E_i = n P_i, i = 1, 2, \dots, k$ . Δηλαδή  $H_0 : E_i = 0_i, i = 1, 2, \dots, k$  με εναλλάκτικη υπόθεση

$H_1 : \text{τουλάχιστον για ένα } i, i=1,2,\dots,k \text{ ισχύει } E_i \neq 0_i$

Ένα κριτήριο (έλεγχος συνάρτηση) για τον έλεγχο της  $H_0$  είναι το κριτήριο  $\chi^2$  που προτάθηκε από τον K. Pearson και ορίζεται από τη σχέση :

$$X^2 = \sum_{i=1}^{\kappa} \frac{(O_i - E_i)^2}{E_i}$$

Η ακριβής κατανομή του κριτηρίου  $X^2$  δεν υπολογίζεται εύκολα και για αυτό καταφεύγουμε στην κατανομή του κριτηρίου για μεγάλα δείγματα. Όταν οι αναμενόμενες συχνότητες  $E_i$  για κάθε κελί  $i$ ,  $i = 1, 2, \dots, \kappa$  είναι  $\geq 5$ , τότε η κατανομή του κριτηρίου  $X^2$  κάτω από την  $H_0$  ( $P_i = \Pi_i$ ) ακολουθεί την κατανομή  $X^2$  με  $\kappa-1$  βαθμούς ελευθερίας. Ένας άλλος περιορισμός που συνήθως γίνεται για τις αναμενόμενες τιμές  $E_i$ , έτσι ώστε το κριτήριο  $X^2$  να ακολουθεί την κατανομή  $X^2$  με  $\kappa-1$  βαθμούς ελευθερίας είναι: Όλες οι αναμενόμενες τιμές  $E_i$  είναι μεγαλύτερες του 1 και το πολύ 20% από αυτές είναι μικρότερες του 5. Γενικά μπορούμε να υιοθετήσουμε τον πρώτο περιορισμό, δηλαδή να απαιτούμε όλα τα  $E_i \geq 5$  [51-53].

Προφανώς μεγάλες διαφορές μεταξύ των παρατηρούμενων συχνοτήτων  $O_i$  και των αναμενόμενων συχνοτήτων  $E_i$  θα οδηγούν στην απόρριψη της  $H_0$ . Ισοδύναμα η μηδενική υπόθεση απορρίπτεται για μεγάλες τιμές του κριτηρίου  $X^2$ . Για δεδομένο επίπεδο σημαντικότητας  $\alpha$ , η κρίσιμη περιοχή αποτελείται από όλες τις τιμές του  $X^2$  που είναι μεγαλύτερες από την ποσότητα  $X_{\kappa-1}^2(\alpha)$ . Δηλαδή  $X^2 > X_{\kappa-1}^2(\alpha)$ .

όπου  $X_{\kappa-1}^2(\alpha)$  το άνω  $\alpha$ -σημείο της κατανομής  $X_{\kappa-1}^2$ , δηλαδή  $P[X_{\kappa-1}^2(a)] = \alpha$ .

Κρίσιμα σημεία  $X_{\nu}^2(\alpha)$  δίνονται από πίνακες για τις διάφορες τιμές των  $\alpha$  και  $\nu$  [51-53].

Στην περιγραφή που κάναμε για τον έλεγχο  $X^2$  μέχρι τώρα υποθέσαμε ότι οι πιθανότητες  $\Pi_i$  κάτω από την μηδενική υπόθεση είναι όλες γνωστές έτσι ώστε να είναι δυνατή η εύρεση των θεωρητικών συχνοτήτων. Υπάρχει όμως και η περίπτωση που μια ή περισσότερες από τις παραμέτρους του υποτιθέμενου μοντέλου/κατανομής να είναι άγνωστες. Π.χ. μπορεί να υποθέσουμε στην  $H_0$  ότι τα δεδομένα ακολουθούν (προέρχονται από) κάποια κατανομή Poisson με παράμετρο  $\lambda$  άγνωστη. Όπως επίσης να υποθέσουμε την κανονική κατανομή με μια από τις παραμέτρους  $\mu$  και  $\sigma^2$  (ή και τις 2) άγνωστες. Η μηδενική υπόθεση θα είναι τώρα της μορφής:

$$H_0 : P_i = \Pi_i(\theta_1, \theta_2, \dots, \theta_r), \quad i=1, 2, \dots, \kappa \quad (r < \kappa).$$

Στην περίπτωση αυτή εκτιμούμε πρώτα τις παραμέτρους  $\theta_1, \theta_2, \dots, \theta_r$  από τα δεδομένα και στη συνέχεια εκτιμούμε τις πιθανότητες  $P_1, P_2, \dots, P_{\kappa}$ . Δηλαδή αν  $\widehat{\theta}_1, \widehat{\theta}_2, \dots, \widehat{\theta}_r$  είναι οι εκτιμήτριες των  $\theta_1, \theta_2, \dots, \theta_r$  υπολογίζουμε στην συνέχεια τις πιθανότητες  $\widehat{P}_i$  όπως προηγουμένως σαν να ήταν γνωστές παράμετροι [51-53]. Οι αναμενόμενες συχνότητες, κάτω από την  $H_0$ , είναι τότε:  $\widehat{E}_i = n\widehat{P}_i$  και το κριτήριο για τον έλεγχο της  $H_0$  γίνεται:

$$\chi^2 = \sum_{i=1}^{\kappa} \frac{(O_i - \hat{E}_i)^2}{\hat{E}_i} = \sum_{i=1}^{\kappa} \frac{(O_i - \nu \hat{P}_i)^2}{\nu \hat{P}_i}$$

Η ασυμπτωτική κατανομή του κριτηρίου  $\chi^2$  είναι η  $\chi^2$  με  $(\kappa-r-1)$  βαθμούς ελευθερίας. Αφαιρούμε δηλαδή επιπλέον το πλήθος των παραμέτρων που εκτιμήσαμε. Η  $H_0$  απορρίπτεται όταν :  $\chi^2 > \chi_{\kappa-r-1}^2(\alpha)$  , όπου

$\kappa$  = αριθμός των κελίων (κατηγοριών)

$r$  = αριθμός των παραμέτρων που εκτιμήσαμε

$\alpha$  = επίπεδο σημαντικότητας

## Κεφάλαιο 3.3 Επιλογή αξιολόγηση και ανάλυση Δεδομένων

Σε αυτό το κεφάλαιο επιλέχθηκαν τα ακόλουθα δεδομένα τα οποία εξήχθησαν από την βάση δεδομένων Seaweb ώστε να μελετηθεί η επίδραση και η συνεισφορά τους στον τύπο του ατυχήματος.

- DWT
- Ηλικία του πλοίου (Age of Ship)
- Χώρα που κτίστηκε το πλοίο (Country Built)
- Ζώνη ατυχήματος (Zone of Accident)
- Είδος του πλοίου (Type of ship)
- Σημαία του πλοίου (Flag of Ship)

Για να αποδειχθεί η εξάρτηση μεταξύ του τύπου ατυχήματος και των ανωτέρω χρησιμοποιήθηκε ο έλεγχος ανεξαρτησίας  $\chi^2$  (Chi square independence test). Με αυτήν την μέθοδο είναι δυνατόν να ελεγχθεί αν υπάρχουν στατιστικά σημαντικές διαφορές στην παρατηρούμενη κατανομή. Αν οι στατιστικές κατανομές εμφανίζουν μεγάλες διαφορές σε σχέση με τις αναμενόμενες μπορεί να θεωρηθεί ότι υπάρχει κάποια συσχέτιση μεταξύ των δύο εξεταζόμενων παραγόντων. Για παράδειγμα αν ο τύπος του ατυχήματος εξαρτάται κατά κάποιο τρόπο από την ηλικία ενός πλοίου.

### Κεφάλαιο 3.3.1 Πληρότητα δεδομένων

Σε αυτό το κεφάλαιο θα γίνει μία προσπάθεια καταγραφής και αξιολόγησης της πληρότητας των δεδομένων που εξήχθησαν από την βάση δεδομένων Sea-web. Οι συνολικές εγγραφές ατυχημάτων για υγραεριοφόρα πλοία ανέρχονται στις 2525 και τα ατυχήματα που συνέβησαν μετά το 1996 είναι 764.

*Πίνακας 1 Αριθμός ατυχημάτων συνολικά και μετά το 1996*

Συνολικά ατυχήματα	2525
Ατυχήματα μετά το 1996	764

*Πίνακας 2 Ελλειπείς εγγραφές αριθμητικά και ποσοστιαία*

	Missing Data	Missing Data (%)	Missing Data after 1996	Missing Data after 1996 (%)
Age of Ship	35	1,39%	12	1,57%
Country Built	1923	76,16%	477	62,43%
Zone	718	28,44%	372	48,69%
Flag	0	0,00%	0	0,00%
Ship type	1923	76,16%	477	62,43%
DWT	4	0,16%	0	0,00%
Type of Accident	537	21,27%	359	46,99%

Ο Πίνακας 2 απεικονίζει τις εγγραφές που απουσιάζουν από την βάση δεδομένων. Τα μεγαλύτερα ποσοστά ελλείψεων τόσο στα συγκεντρωτικά στοιχεία όσο και στα στοιχεία μετά το 1996 εμφανίζονται στις κατηγορίες “Country Built” και “Ship Type”. Παράλληλα στις κατηγορίες “Flag”, “Age of Ship” και “DWT” παρουσιάζονται μηδενικές ή σχεδόν μηδενικές ελλείψεις. Σημαντικό είναι το γεγονός ότι από τα 764 ατυχήματα που συνέβησαν μετά το 1996 τα 359 δεν έχουν επαρκή στοιχεία όσον αφορά τον τύπο του ατυχήματος (46.99%). Αυτό το στοιχείο περιορίζει σημαντικά το στατιστικό δείγμα που έχουμε στην διάθεση μας.



## Κεφάλαιο 3.3.2 Εξάρτηση είδους ατυχήματος – δεδομένων

Ο έλεγχος με χρήση της μεθόδου  $\chi^2$  έχει πολλές παραλλαγές και διαφοροποιήσεις. Στην συγκεκριμένη περίπτωση τα βήματα της μεθοδολογίας τα οποία ακολουθήθηκαν για τον έλεγχο της στατιστικής κατανομής των δεδομένων είναι τα ακόλουθα:

1. Δημιουργία πίνακα (K) 2 διαστάσεων ποσοτικής καταγραφής των ατυχημάτων ως προς τον τύπο του ατυχήματος και ενός εκ των παραγόντων.

2. Υπολογισμός βαθμών ελευθερίας βάσει του ακόλουθου τύπου:

$$Df=(n-1)*(m-1)$$

όπου n= Αριθμός στηλών m= Αριθμός γραμμών

3. Δημιουργία πίνακα E διαστάσεων n,m βάσει του ακόλουθου τύπου:

$$E_{m,n}=(\text{Άθροισμα των ατυχημάτων της στήλης } n) * (\text{Άθροισμα των ατυχημάτων της γραμμής } m) / (\text{Άθροισμα των ολικών ατυχημάτων})$$

4. Δημιουργία πίνακα X βάσει του ακόλουθου τύπου:

$$X_{m,n} = (K_{m,n} * E_{m,n})^2 / E_{m,n}$$

5. Το  $\chi^2$  το οποίο μας ενδιαφέρει προκύπτει ως το άθροισμα όλων των επιμέρους X.

6. Το P<sub>i</sub> υπολογίζεται μέσω των  $\chi^2$  και Df.

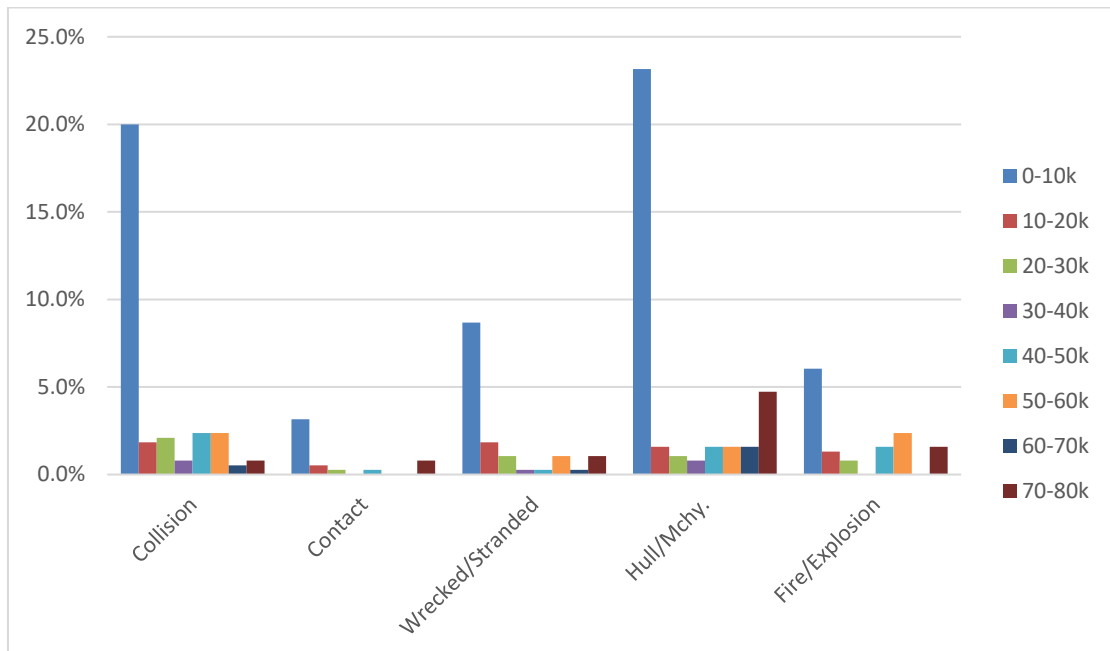
Το P<sub>i</sub> εκφράζει την πιθανότητα εμφάνισης του test statistic που υπολογίζεται από τα δεδομένα στη θεωρητική κατανομή. Για τιμές του P<sub>i</sub> μικρότερες του 0.05 (5%) μπορούμε ασφαλώς να υποθέσουμε ότι η κατανομή των δεδομένων μας δεν ακολουθεί την θεωρητική κατανομή. Ακολουθούν οι πίνακες και τα αποτελέσματα για τους 6 παράγοντες που υποθέσαμε ότι επηρεάζουν τον τύπο του ατυχήματος.

Στατιστική Ανάλυση Ναυτικών Ατυχημάτων σε πλοία LNG-LPG  
και Ανάπτυξη Αλγορίθμου Εξόρυξης Δεδομένων

**-DWT**

Πίνακας 3 Κατανομή ατυχημάτων συναρτήσει του DWT

DWT	Collision		Contact		Wrecked/Stranded		Hull/Mchy.		Fire/Explosion		Total R	
0-10k	20,00%	76	3,16%	12	8,68%	33	23,16%	88	6,05%	23	61,05%	232
10-20k	1,84%	7	0,53%	2	1,84%	7	1,58%	6	1,32%	5	7,11%	27
20-30k	2,11%	8	0,26%	1	1,05%	4	1,05%	4	0,79%	3	5,26%	20
30-40k	0,79%	3	0,00%	0	0,26%	1	0,79%	3	0,00%	0	1,84%	7
40-50k	2,37%	9	0,26%	1	0,26%	1	1,58%	6	1,58%	6	6,05%	23
50-60k	2,37%	9	0,00%	0	1,05%	4	1,58%	6	2,37%	9	7,37%	28
60-70k	0,53%	2	0,00%	0	0,26%	1	1,58%	6	0,00%	0	2,37%	9
70-80k	0,79%	3	0,79%	3	1,05%	4	4,74%	18	1,58%	6	8,95%	34
Total C	30,79%	117	5,00%	19	14,47%	55	36,05%	137	13,68%	52	100,00%	380



Σχήμα 1 Κατανομή ατυχημάτων συναρτήσει του DWT

Η εφαρμογή του στατιστικού test  $\chi^2$  για  $Df = 28$  και  $\chi^2 = 41.29$  έδειξε ότι:

$$P(Df=28, \chi^2 > 41.29) = 0.07 (7\%) > 0.05$$

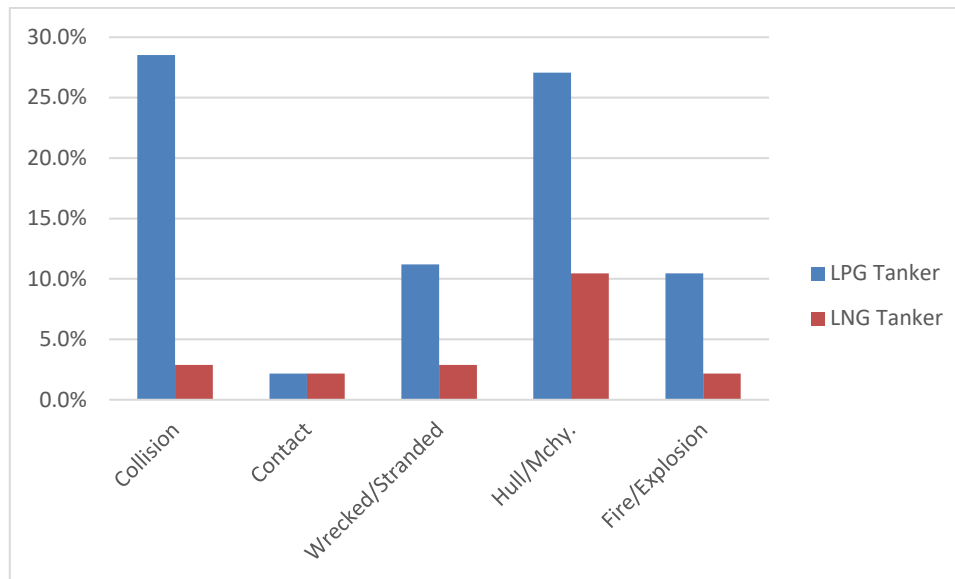
Άρα βάση αυτού του στατιστικού test το DWT δεν μπορεί να χρησιμοποιηθεί για την εξαγωγή διαγραμμάτων και σχέσεων αιτίου-αποτελέσματος σε σχέση με τον τύπο του ατυχήματος με αξιοπιστία.

Στατιστική Ανάλυση Ναυτικών Ατυχημάτων σε πλοία LNG-LPG  
και Ανάπτυξη Αλγορίθμου Εξόρυξης Δεδομένων

**-Είδος πλοίου**

Πίνακας 4 Κατανομή ατυχημάτων συναρτήσει του Είδους του πλοίου

Type of Ship	Collision		Contact		Wrecked/Stranded		Hull/Mchy.		Fire/Explosion		Total R	
LPG Tanker	28,52%	79	2,17%	6	11,19%	31	27,08%	75	10,47%	29	79,42%	220
LNG Tanker	2,89%	8	2,17%	6	2,89%	8	10,47%	29	2,17%	6	20,58%	57
Total C	31,41%	87	4,33%	12	14,08%	39	37,55%	104	12,64%	35	100,00%	277



Σχήμα 2 Κατανομή ατυχημάτων συναρτήσει του Είδους του πλοίου

Η εφαρμογή του στατιστικού test  $\chi^2$  για  $Df = 4$  και  $\chi^2 = 16.90$  έδειξε ότι:

$$P(Df=4, \chi^2 > 16.90) = 0.02 (2\%) < 0.05$$

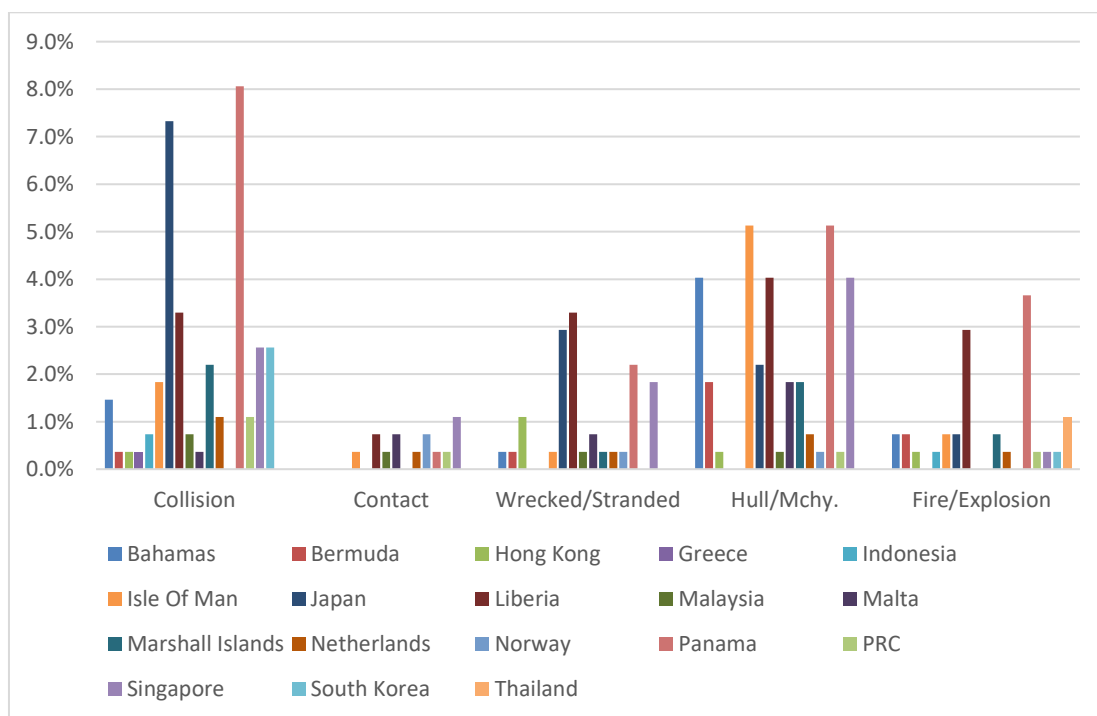
Άρα βάση αυτού του στατιστικού test το είδος του πλοίου μπορεί να χρησιμοποιηθεί για την εξαγωγή διαγραμμάτων και σχέσεων αιτίου-αποτελέσματος σε σχέση με τον τύπο του ατυχήματος με αξιοπιστία.

Στατιστική Ανάλυση Ναυτικών Ατυχημάτων σε πλοία LNG-LPG  
και Ανάπτυξη Αλγορίθμου Εξόρυξης Δεδομένων

**-Σημαία του πλοίου**

Πίνακας 5 Κατανομή ατυχημάτων συναρτήσει της σημαίας του πλοίου

Flag	Collision	Contact	Wrecked/Stranded	Hull/Mchy.	Fire/Explosion	Total R						
Bahamas	1,47%	4	0,00%	0	0,37%	1	4,03%	11	0,73%	2	6,59%	18
Bermuda	0,37%	1	0,00%	0	0,37%	1	1,83%	5	0,73%	2	3,30%	9
China Hong Kong	0,37%	1	0,00%	0	1,10%	3	0,37%	1	0,37%	1	2,20%	6
Greece	0,37%	1	0,00%	0	0,00%	0	0,00%	0	0,00%	0	0,37%	1
Indonesia	0,73%	2	0,00%	0	0,00%	0	0,00%	0	0,37%	1	1,10%	3
Isle Of Man	1,83%	5	0,37%	1	0,37%	1	5,13%	14	0,73%	2	8,42%	23
Japan	7,33%	20	0,00%	0	2,93%	8	2,20%	6	0,73%	2	13,19%	36
Liberia	3,30%	9	0,73%	2	3,30%	9	4,03%	11	2,93%	8	14,29%	39
Malaysia	0,73%	2	0,37%	1	0,37%	1	0,37%	1	0,00%	0	1,83%	5
Malta	0,37%	1	0,73%	2	0,73%	2	1,83%	5	0,00%	0	3,66%	10
Marshall Islands	2,20%	6	0,00%	0	0,37%	1	1,83%	5	0,73%	2	5,13%	14
Netherlands	1,10%	3	0,37%	1	0,37%	1	0,73%	2	0,37%	1	2,93%	8
Norway	0,00%	0	0,73%	2	0,37%	1	0,37%	1	0,00%	0	1,47%	4
Panama	8,06%	22	0,37%	1	2,20%	6	5,13%	14	3,66%	10	19,41%	53
People's Republic Of China	1,10%	3	0,37%	1	0,00%	0	0,37%	1	0,37%	1	2,20%	6
Singapore	2,56%	7	1,10%	3	1,83%	5	4,03%	11	0,37%	1	9,89%	27
South Korea	2,56%	7	0,00%	0	0,00%	0	0,00%	0	0,37%	1	2,93%	8
Thailand	0,00%	0	0,00%	0	0,00%	0	0,00%	0	1,10%	3	1,10%	3
Total C	34,43%	94	5,13%	14	14,65%	40	32,23%	88	13,55%	37	100,00%	273



Σχήμα 3 Κατανομή ατυχημάτων συναρτήσει της σημαίας του πλοίου

Στατιστική Ανάλυση Ναυτικών Ατυχημάτων σε πλοία LNG-LPG  
και Ανάπτυξη Αλγορίθμου Εξόρυξης Δεδομένων

Η εφαρμογή του στατιστικού test  $\chi^2$  για  $Df = 72$  και  $\chi^2 = 124.21$  έδειξε ότι:

$$P(Df=72, \chi^2 > 124.21) = 0.0001 (0.01\%) < 0.05$$

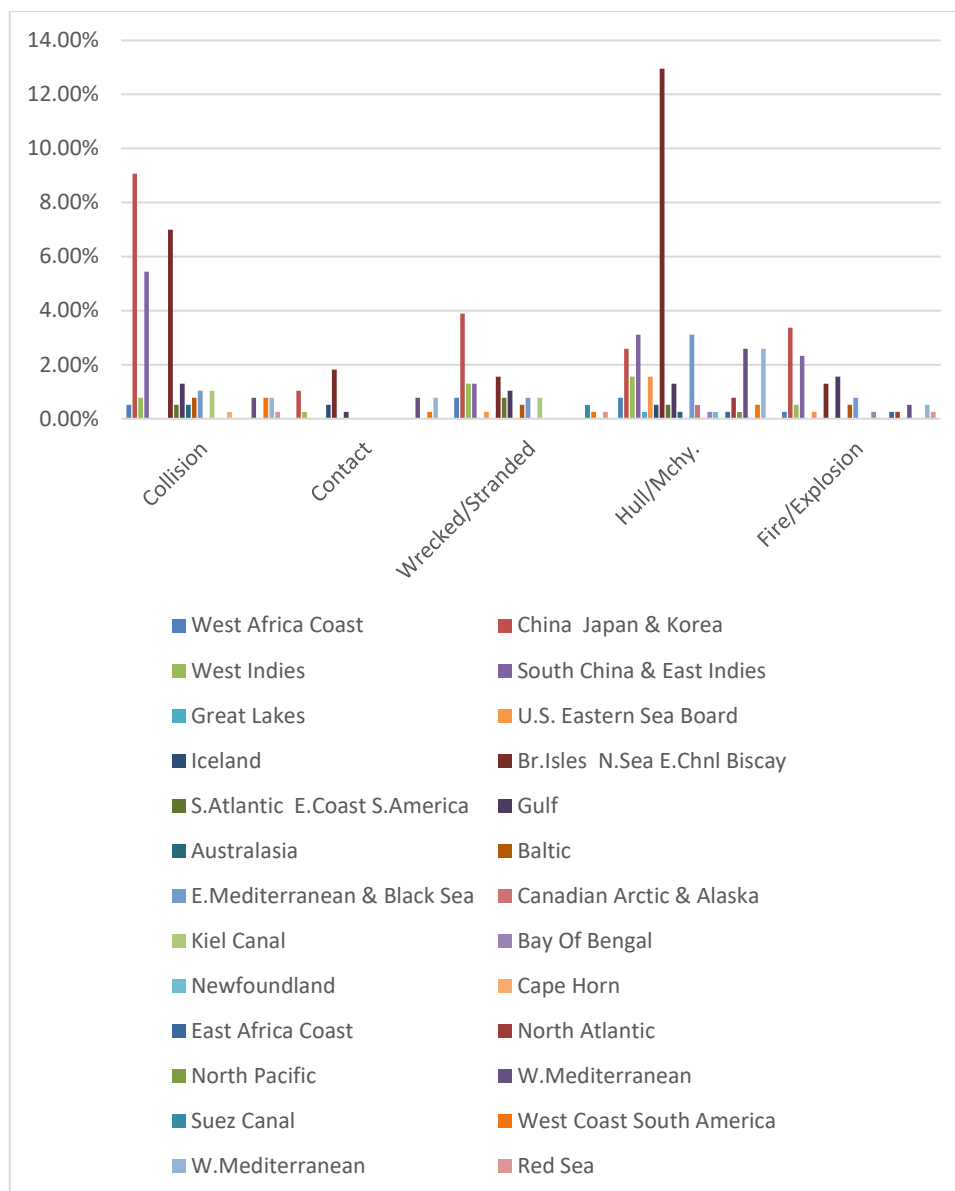
Άρα βάση αυτού του στατιστικού test η σημαία του πλοίου μπορεί να χρησιμοποιηθεί για την εξαγωγή διαγραμμάτων και σχέσεων αιτίου-αποτελέσματος σε σχέση με τον τύπο του ατυχήματος με αξιοπιστία.

### -Ζώνη ατυχήματος

Πίνακας 6 Κατανομή ατυχημάτων συναρτήσει της ζώνης ατυχήματος

Zone	Collision	Contact	Wrecked/Stranded	Hull/Mchy.	Fire/Explosion	Total R
West Africa Coast	0,52%	2	0,00%	0	0,78%	3
China Japan & Korea	9,07%	35	1,04%	4	3,89%	15
West Indies	0,78%	3	0,26%	1	1,30%	5
South China & East Indies	5,44%	21	0,00%	0	1,30%	5
Great Lakes	0,00%	0	0,00%	0	0,00%	0
U.S. Eastern Sea Board	0,00%	0	0,00%	0	0,26%	1
Iceland	0,00%	0	0,52%	2	0,00%	0
Br.Isles N.Sea E.Chnl Biscay	6,99%	27	1,81%	7	1,55%	6
S.Atlantic E.Coast S.America	0,52%	2	0,00%	0	0,78%	3
Gulf	1,30%	5	0,26%	1	1,04%	4
Australasia	0,52%	2	0,00%	0	0,00%	0
Baltic	0,78%	3	0,00%	0	0,52%	2
E.Mediterranean & Black Sea	1,04%	4	0,00%	0	0,78%	3
Canadian Arctic & Alaska	0,00%	0	0,00%	0	0,00%	0
Kiel Canal	1,04%	4	0,00%	0	0,78%	3
Bay Of Bengal	0,00%	0	0,00%	0	0,00%	0
Newfoundland	0,00%	0	0,00%	0	0,26%	1
Cape Horn	0,26%	1	0,00%	0	0,00%	0
East Africa Coast	0,00%	0	0,00%	0	0,00%	0
North Atlantic	0,00%	0	0,00%	0	0,78%	3
North Pacific	0,00%	0	0,00%	0	0,26%	1
W.Mediterranean	0,78%	3	0,78%	3	0,00%	0
Suez Canal	0,00%	0	0,00%	0	0,52%	2
West Coast South America	0,78%	3	0,26%	1	0,26%	1
W.Mediterranean	0,78%	3	0,78%	3	0,00%	0
Red Sea	0,26%	1	0,00%	0	0,26%	1
Total C	30,83%	119	5,70%	22	13,99%	54

Στατιστική Ανάλυση Ναυτικών Ατυχημάτων σε πλοία LNG-LPG  
και Ανάπτυξη Αλγορίθμου Εξόρυξης Δεδομένων



Σχήμα 4 Κατανομή ατυχημάτων συναρτήσει της ζώνης ατυχήματος

Η εφαρμογή του στατιστικού test  $\chi^2$  για  $Df = 100$  και  $\chi^2 = 164.64$  έδειξε ότι:

$$P(Df=100, \chi^2 > 164.64) = 0.0001 (0.01\%) < 0.05$$

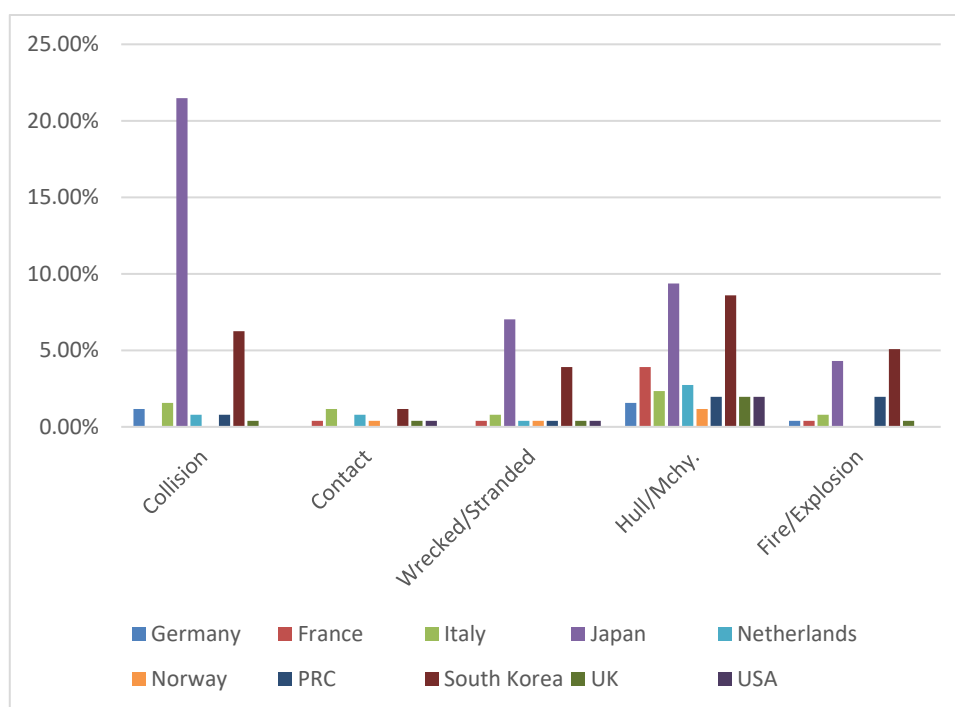
Άρα βάση αυτού του στατιστικού test η ζώνη που συνέβη το ατύχημα μπορεί να χρησιμοποιηθεί για την εξαγωγή διαγραμμάτων και σχέσεων αιτίου-αποτελέσματος σε σχέση με τον τύπο του ατυχήματος με αξιοπιστία.

Στατιστική Ανάλυση Ναυτικών Ατυχημάτων σε πλοία LNG-LPG  
και Ανάπτυξη Αλγορίθμου Εξόρυξης Δεδομένων

**-Χώρα που κτίστηκε το πλοίο**

Πίνακας 7 Κατανομή ατυχημάτων συναρτήσει της χώρας που κτίστηκε το πλοίο

Country Built	Collision	Contact	Wrecked/Stranded	Hull/Mchy.	Fire/Explosion	Total R
Federal Republic Of Germany	1,17%	3	0,00%	0	0,00%	3,13%
France	0,00%	0	0,39%	1	0,39%	5,08%
Italy	1,56%	4	1,17%	3	0,78%	6,64%
Japan	21,48%	55	0,00%	0	7,03%	42,19%
Netherlands	0,78%	2	0,78%	2	0,39%	4,69%
Norway	0,00%	0	0,39%	1	0,39%	1,95%
People's Republic Of China	0,78%	2	0,00%	0	0,39%	5,08%
South Korea	6,25%	16	1,17%	3	3,91%	25,00%
United Kingdom	0,39%	1	0,39%	1	0,39%	3,52%
United States Of America	0,00%	0	0,39%	1	1,95%	2,73%
Total C	32,42%	83	4,69%	12	14,06%	36
						35,55%
						91
						13,28%
						34
						100,00%
						256



Σχήμα 5 Κατανομή ατυχημάτων συναρτήσει της χώρας που κτίστηκε το πλοίο

Η εφαρμογή του στατιστικού test  $\chi^2$  για  $Df = 36$  και  $\chi^2 = 79.43$  έδειξε ότι:

$$P(Df=36, \chi^2 > 79.43) = 0.00001 (0.001\%) < 0.05$$

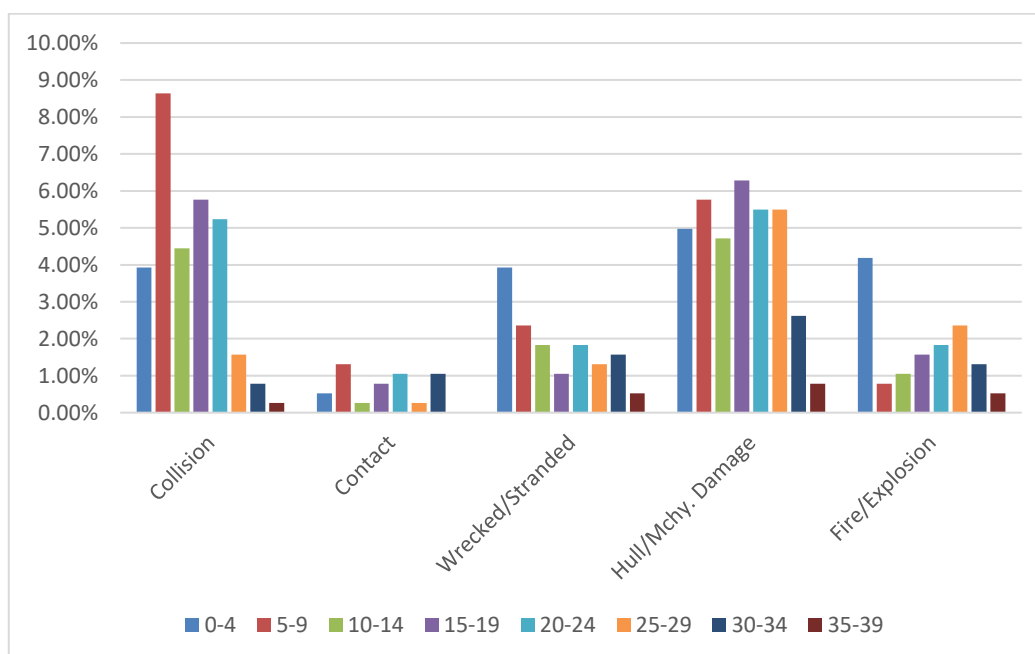
Άρα βάση αυτού του στατιστικού test η χώρα στην οποία χτίστηκε το πλοίο μπορεί να χρησιμοποιηθεί για την εξαγωγή διαγραμμάτων και σχέσεων αιτίου-αποτελέσματος σε σχέση με τον τύπο του ατυχήματος με αξιοπιστία.

Στατιστική Ανάλυση Ναυτικών Ατυχημάτων σε πλοία LNG-LPG  
και Ανάπτυξη Αλγορίθμου Εξόρυξης Δεδομένων

**-Ηλικία του πλοίου**

Πίνακας 8 Κατανομή ατυχημάτων συναρτήσει της ηλικίας του πλοίου

Age of ship	Collision		Contact		Wrecked/Stranded		Hull/Mchy.		Fire/Explosion		Total R	
0-4	3,93%	15	0,52%	2	3,93%	15	4,97%	19	4,19%	16	17,54%	67
5-9	8,64%	33	1,31%	5	2,36%	9	5,76%	22	0,79%	3	18,85%	72
10-14	4,45%	17	0,26%	1	1,83%	7	4,71%	18	1,05%	4	12,30%	47
15-19	5,76%	22	0,79%	3	1,05%	4	6,28%	24	1,57%	6	15,45%	59
20-24	5,24%	20	1,05%	4	1,83%	7	5,50%	21	1,83%	7	15,45%	59
25-29	1,57%	6	0,26%	1	1,31%	5	5,50%	21	2,36%	9	10,99%	42
30-34	0,79%	3	1,05%	4	1,57%	6	2,62%	10	1,31%	5	7,33%	28
35-39	0,26%	1	0,00%	0	0,52%	2	0,79%	3	0,52%	2	2,09%	8
Total C	30,63%	117	5,24%	20	14,40%	55	36,13%	138	13,61%	52	100,00%	382



Σχήμα 6 Κατανομή ατυχημάτων συναρτήσει της ηλικίας του πλοίου

Η εφαρμογή του στατιστικού test  $\chi^2$  για  $Df = 28$  και  $\chi^2 = 50.66$  έδειξε ότι:

$$P(Df=28, \chi^2 > 50.66) = 0.005 (0.5\%) < 0.05$$

Άρα βάση αυτού του στατιστικού test η ηλικία του πλοίου όταν συνέβη το ατύχημα μπορεί να χρησιμοποιηθεί για την εξαγωγή διαγραμμάτων και σχέσεων αιτίου-αποτελέσματος σε σχέση με τον τύπο του ατυχήματος με αξιοπιστία.



## -Τιμές Pi

Πίνακας 9 Συνολικές τιμές Pi και  $\chi^2$  για όλα τα δεδομένα

	DWT	Type of Ship	Flag of Ship	Zone of Accident	Country Built	Age of Ship
$\chi^2 =$	41,29	16,90	124,21	164,64	79,43	50,66
Pi =	7,00%	2,00%	0,01%	0,01%	0,001%	0,50%

Για τιμές του Pi μικρότερες ή ίσες του 5% μπορεί να θεωρηθεί με ικανοποιητική ακρίβεια ότι ο τύπος του ατυχήματος εξαρτάται από το εκάστοτε δεδομένο. Για τις προκύπτουσες τιμές Pi έχουμε ικανοποιητική ακρίβεια για όλα τα δεδομένα πλην του DWT. Για το DWT δεν είναι δυνατόν με τα παρών στοιχεία χρησιμοποιώντας τον έλεγχο του  $\chi^2$  να προσδιορίσουμε με σιγουριά αν επηρεάζει τον τύπο του ατυχήματος. Το ποσοστό της τάξης του 7% το οποίο και προέκυψε δεν μας δίνει την δυνατότητα να θεωρήσουμε ότι είναι ούτε εξαρτημένο ούτε ανεξάρτητο σε σχέση με τον τύπο ατυχήματος. Συνεπώς στην διαδικασία της μελέτης των ατυχημάτων δεν θα ληφθεί υπόψη η επίδραση του DWT.

Η σελίδα αυτή είναι σκόπιμα κενή.

## ΚΕΦΑΛΑΙΟ 4. Στατιστική Ανάλυση Ατυχημάτων

Στο συγκεκριμένο κεφάλαιο θα πραγματοποιηθεί βάσει των στοιχείων της βάσης δεδομένων Sea-web στατιστική ανάλυση των ατυχημάτων σε πλοία LNG-LPG για τα έτη 1997-2016. Στόχος της συγκεκριμένης ανάλυσης είναι ο υπολογισμός του βαθμού στον οποίο οι διάφοροι παράγοντες επηρεάζουν την ύπαρξη ή μη ενός ατυχήματος ή ενός συγκεκριμένου τύπου ατυχήματος και η εξαγωγή συχνοτήτων ατυχημάτων. Έχοντας αυτόν τον στόχο ακολουθεί η ανάλυση των ατυχημάτων βάσει των ακόλουθων παραμέτρων:

- Τον τύπο του πλοίου
- Την σημαία του πλοίου
- Την περιοχή που συνέβη το ατύχημα
- Την ηλικία του πλοίου όταν συνέβη το ατύχημα
- Την χώρα κατασκευής

Έχοντας μελετήσει την στατιστική κατανομή των δεδομένων συναρτήσει των ανωτέρω παραγόντων μπορούμε να υποθέσουμε με ικανοποιητική ακρίβεια ότι αυτοί επηρεάζουν τον τύπο ατυχήματος και συνεπώς θα επιχειρήσουμε να εξάγουμε διαγράμματα κατανομής ατυχημάτων και να συμπεράνουμε για κάθε ένα είδος ατυχήματος αλλά και συγκεντρωτικά ποια πλοία είναι περισσότερο πιθανόν να εμπλακούν σε κάποιο ατύχημα.

Επίσης μελετήσαμε την επίδραση του DWT στον τύπο ατυχήματος και δεν ήταν δυνατή η εξαγωγή τελικού συμπεράσματος για την εξάρτηση ή ανεξαρτησία του τύπου ατυχήματος από το DWT. Συνεπώς δεν θα γίνει καμία προσπάθεια να εξαχθούν διαγράμματα συναρτήσει του DWT ή να βρούμε σχέσεις αλληλεπίδρασης με τον τύπο ατυχήματος.

Αρχικά θα μελετήσουμε για τα μεγέθη τα οποία είναι δυνατόν (Type of Ship, Flag of Ship, Country Built) την στατιστική κατανομή ολόκληρου του στόλου. Έτσι θα μας δοθεί η δυνατότητα να καθορίσουμε σε τι βαθμό η στατιστική κατανομή των ατυχημάτων ακολουθεί την στατιστική κατανομή του στόλου και σε τι βαθμό αποκλίνει. Αν η κατανομή των ατυχημάτων ανά είδος ατυχήματος αποκλίνει σε σημαντικό βαθμό από την κατανομή του στόλου τότε άμεσα συμπεραίνουμε ότι συγκεκριμένα πλοία είναι πιο επιρρεπή σε συγκεκριμένα ατυχήματα.

Για τα μεγέθη “Zone of Accident” και “Age of Ship when accident happened” δεν είναι δυνατόν να υπάρξει μία κατανομή στόλου όπως στις ανωτέρω ποσότητες. Συνεπώς η κατανομή των ατυχημάτων συναρτήσει αυτών των δύο μεγεθών ανά κατηγορία ατυχήματος θα συγκριθεί με την συνολική κατανομή των ατυχημάτων και όχι με την κατανομή του στόλου.

Κάθε τύπος ατυχήματος θα μελετηθεί ξεχωριστά και τελικά θα εξαχθούν γενικά συμπεράσματα. Τέλος θα εξαχθούν οι συχνότητες των ατυχημάτων καθώς και οι συχνότητες των ατυχημάτων συναρτήσει της ηλικίας του πλοίου.

## Κεφάλαιο 4.1 Στατιστική ανάλυση ατυχημάτων υγραεριοφόρων πλοίων

Σε αυτό το κεφάλαιο αρχικά θα γίνει μία ανάλυση της κατανομής του στόλου βάσει των προαναφερθέντων παραγόντων. Ο στόλος και τα στοιχεία του τα οποία θα αναλυθούν είναι αυτά τα οποία εξήχθησαν από την βάση δεδομένων Sea-web. Συνολικά ο στόλος αποτελείται από 2154 καταγεγραμμένα πλοία. Ως πλοία του στόλου χρησιμοποιήθηκαν όλα τα πλοία τα οποία βρίσκονται εν ενεργεία κατά την εξαγωγή των στοιχείων (2016).

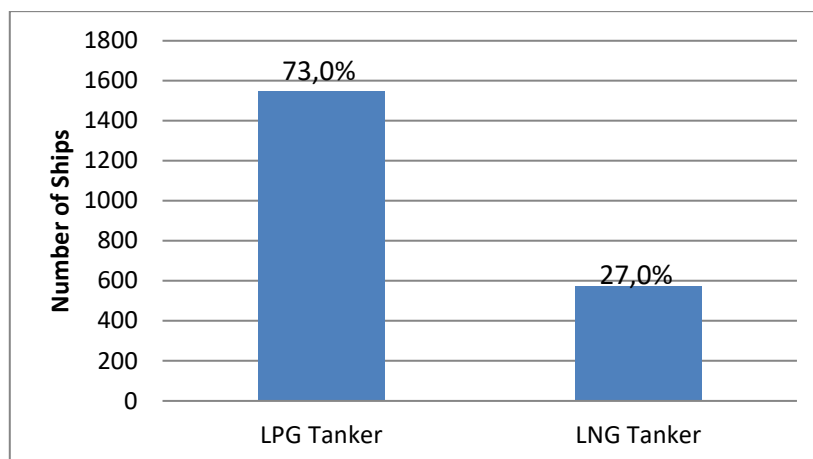
Στη συνέχεια θα γίνει η ανάλυση των ατυχημάτων των πλοίων συνολικά και ανά κατηγορία ατυχήματος τα οποία συνέβησαν μετά το 1996. Στον Πίνακα 10 Κατανομή ατυχημάτων ανά κατηγορία ατυχήματος εμφανίζεται η κατανομή των ατυχημάτων που έλαβαν χώρα μετά το 1996 ανά κατηγορία ατυχήματος.

*Πίνακας 10 Κατανομή ατυχημάτων ανά κατηγορία ατυχήματος*

Type of Accident	Accidents	Percentage
Collision	121	29,88%
Wrecked/Stranded	58	14,32%
Hull/Mchy. Damage	143	35,31%
Fire/Explosion	53	13,09%
Contact	20	4,94%
Foundered	10	2,47%
War Loss/Hostilities	0	0,00%

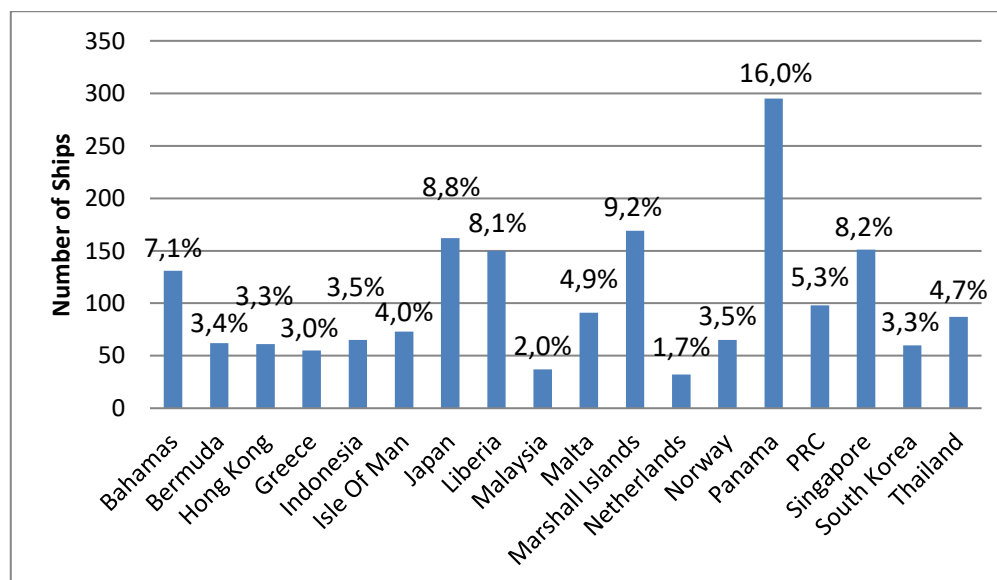
Οι κατηγορίες των “Contact”, “Foundered”, “War Loss/Hostilities” θα εξαιρεθούν της ανάλυσης καθώς παρουσιάζουν πολύ μικρό αριθμό συνολικών ατυχημάτων και δεν είναι δυνατόν να εξαχθούν αξιόλογα στοιχεία.

## Κεφάλαιο 4.1.1 Κατανομή στόλου



Σχήμα 7 Κατανομή του στόλου συναρτήσει του Ship Type

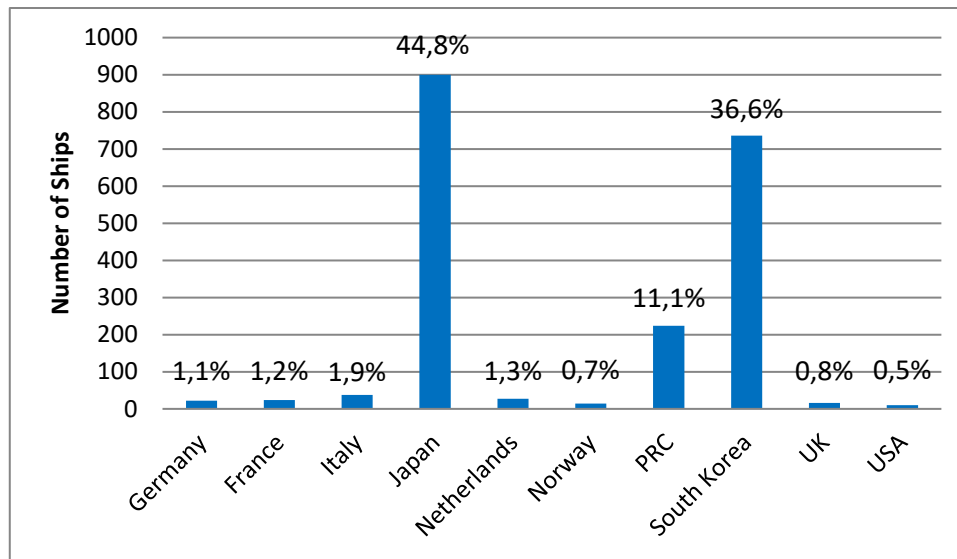
Στο Σχήμα 7 παρουσιάζεται η κατανομή του στόλου συναρτήσει του τύπου του πλοίου. Παρατηρείται ότι ο αριθμός των LPG Tanker (73%) είναι περίπου τριπλάσιος από τον αριθμό των LNG(27%). Συγκεκριμένα ο αριθμός των καταγεγραμμένων πλοίων LPG ανέρχεται στα 1549 ενώ αντίστοιχα τα πλοία μεταφοράς LNG είναι μόλις 574.



Σχήμα 8 Κατανομή του στόλου συναρτήσει της σημαίας

Στατιστική Ανάλυση Ναυτικών Ατυχημάτων σε πλοία LNG-LPG  
και Ανάπτυξη Αλγορίθμου Εξόρυξης Δεδομένων

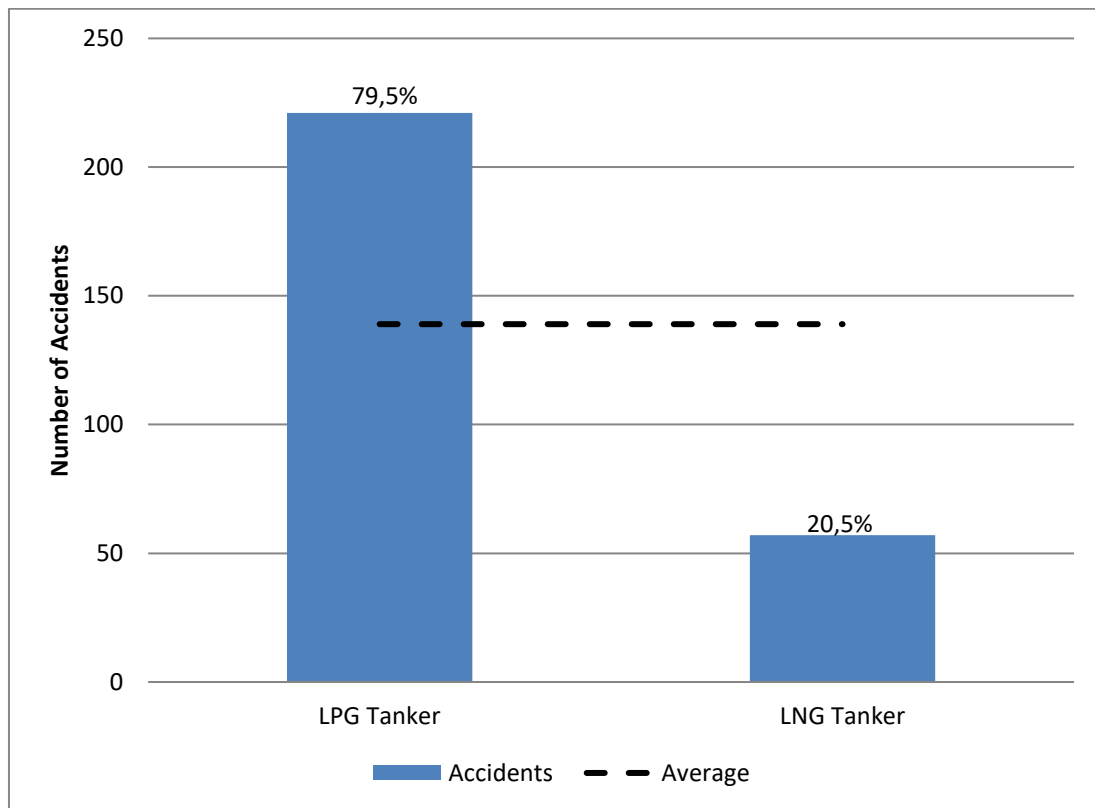
Στο Σχήμα 8 παρουσιάζεται η κατανομή του στόλου συναρτήσει της σημαίας του πλοίου. Αναλυτικότερα η σημαία με τα περισσότερα εγγεγραμμένα πλοία είναι αυτή του Παναμά (16%) με 295 εγγεγραμμένα πλοία ενώ ακολουθούν οι Νήσοι Μάρσαλ (9.2%), η Ιαπωνία (8.8%), η Σιγκαπούρη (8.2%), η Λιβερία (8.1%), και οι Μπαχάμες (7.1%) με 120-180 πλοία έκαστος. Όλες οι υπόλοιπες σημαίες έχουν λιγότερα από 100 εγγεγραμμένα πλοία.



Σχήμα 9 Κατανομή του στόλου συναρτήσει του Country Built

Στο Σχήμα 9 παρουσιάζεται η κατανομή του στόλου συναρτήσει της χώρας κατασκευής του πλοίου. Από το διάγραμμα είναι ξεκάθαρο ότι οι 3 χώρες που αναλαμβάνουν κυρίως την κατασκευή υγραεριοφόρων πλοίων είναι η Ιαπωνία(44.8%) με 900 ναυπηγημένα πλοία, η Νότια Κορέα(36.6%) με 736 ναυπηγημένα πλοία και η Κίνα(11.1%) με 224 ναυπηγημένα πλοία. Οι τρεις αυτές χώρες συνολικά έχουν ναυπηγήσει το 92.5% του ενεργού στόλου.

## Κεφάλαιο 4.1.2 Συγκεντρωτική κατανομή ατυχημάτων

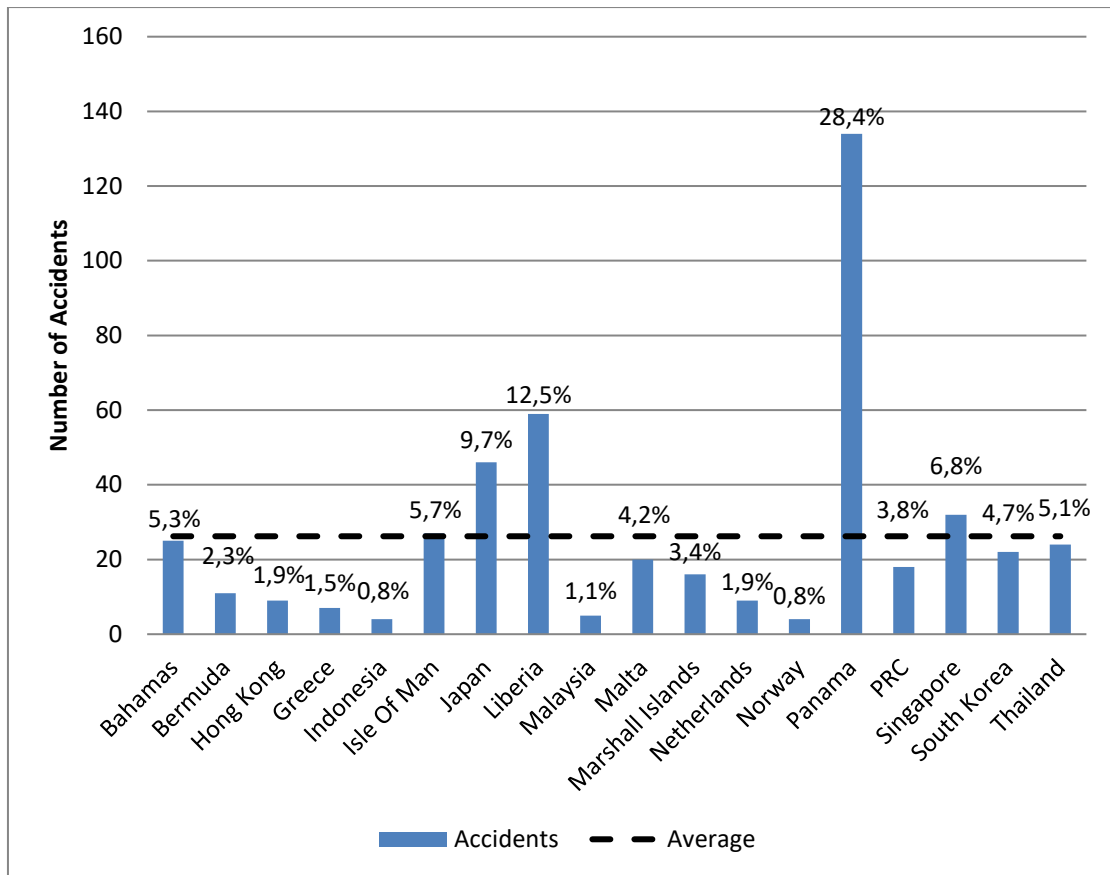


Σχήμα 10 Συγκεντρωτική κατανομή ατυχημάτων συναρτήσει του Ship Type

Στο Σχήμα 10 παρουσιάζεται η συγκεντρωτική κατανομή των ατυχημάτων συναρτήσει του τύπου του πλοίου. Ο μέσος όρος ατυχημάτων ανά τύπο πλοίου είναι 139 ατυχήματα (διακεκομμένη γραμμή). Τα ατυχήματα τα οποία έχουν συμβεί σε LPG Tankers αποτελούν το 79.5% ενώ τα ατυχήματα που έχουν συμβεί σε LNG Tankers αποτελούν το 20.5%. Δεδομένου ότι ο στόλος των LPG Tankers είναι το 73% του συνολικού στόλου είναι ξεκάθαρο ότι η συχνότητα ύπαρξης ατυχημάτων στα πλοία μεταφοράς LPG είναι ελαφρώς μεγαλύτερη συγκριτικά με τα πλοία LNG.



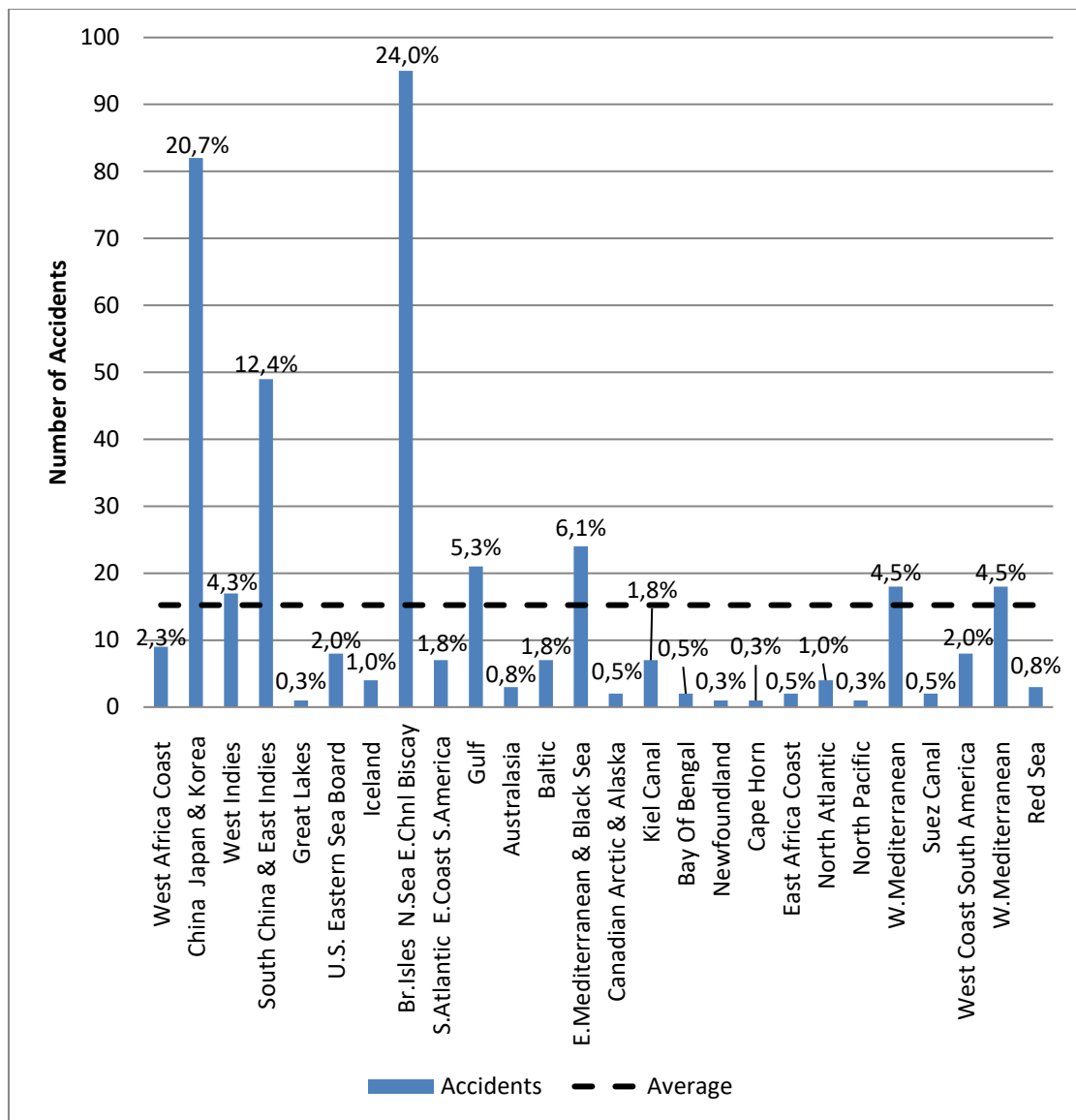
Στατιστική Ανάλυση Ναυτικών Ατυχημάτων σε πλοία LNG-LPG  
και Ανάπτυξη Αλγορίθμου Εξόρυξης Δεδομένων



Σχήμα 11 Συγκεντρωτική κατανομή ατυχημάτων συναρτήσει του Flag

Στο Σχήμα 11 παρουσιάζεται η συγκεντρωτική κατανομή των ατυχημάτων συναρτήσει της σημαίας του πλοίου. Ο μέσος όρος ατυχημάτων ανά σημαία είναι 26.22 ατυχήματα. Η σημαία του Παναμά εμφανίζει τα περισσότερα ατυχήματα (28.4%). Ακολουθεί η σημαία της Λιβερίας με 12.5% των ατυχημάτων και της Ιαπωνίας με 9.7%. Τέλος ακολουθούν οι σημαίες της Ταϊλάνδης, της Νότιας Κορέας, Της Σιγκαπούρης, της Κίνας, των Μπαχαμών και του Isle of Man με 20-30 (4.5%-7%) ατυχήματα έκαστος. Πέραν της σημαίας της Λιβερίας στην οποία υπάγεται μόλις το 8.1% του στόλου και το 12.5% των ατυχημάτων και της σημαίας του Παναμά στην οποία υπάγεται το 16% του στόλου και το 28.4% των ατυχημάτων οι υπόλοιπες σημαίες δεν εμφανίζουν ιδιαίτερες διαφοροποιήσεις μεταξύ των ποσοστών στόλου και ατυχημάτων.

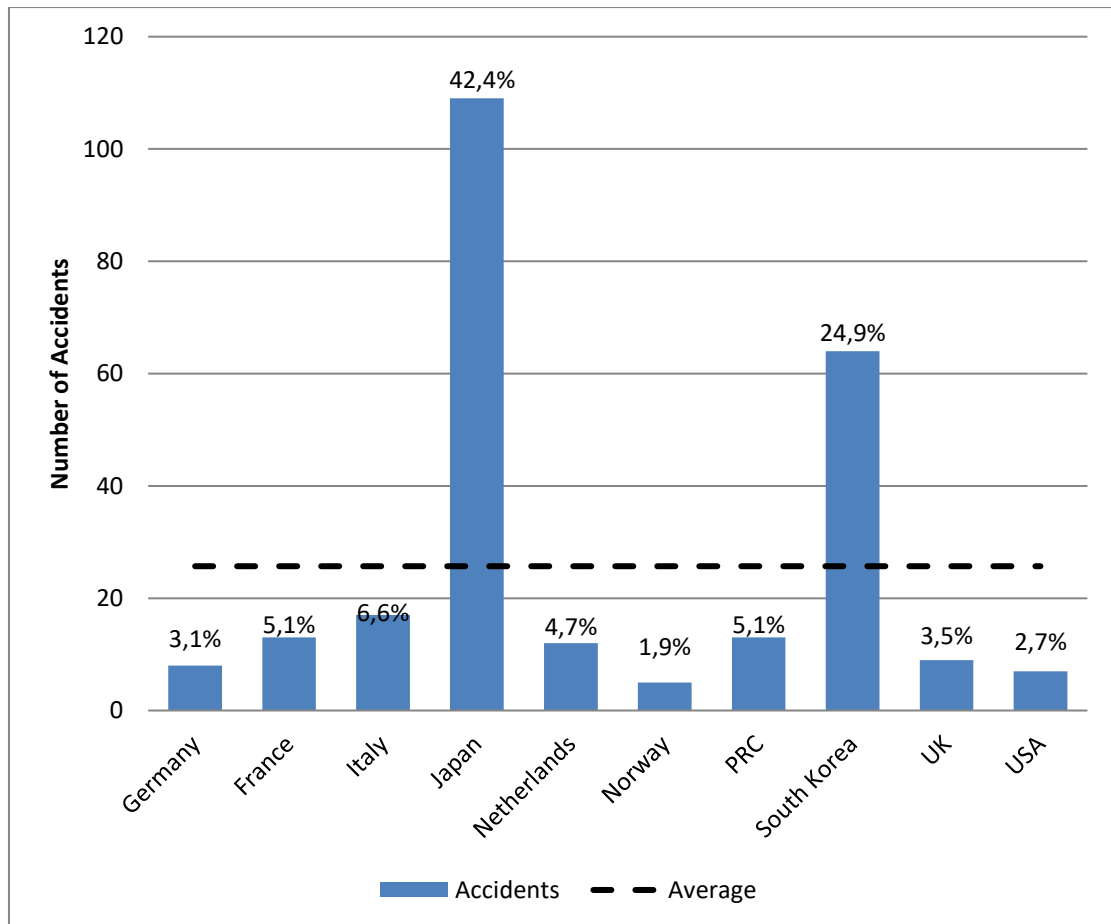
Στατιστική Ανάλυση Ναυτικών Ατυχημάτων σε πλοία LNG-LPG  
και Ανάπτυξη Αλγορίθμου Εξόρυξης Δεδομένων



Σχήμα 12 Συγκεντρωτική κατανομή ατυχημάτων συναρτήσει του Zone

Στο Σχήμα 12 παρουσιάζεται η συγκεντρωτική κατανομή των ατυχημάτων συναρτήσει της ζώνης του ατυχήματος. Ο μέσος όρος ατυχημάτων ανά ζώνη είναι 15.23 ατυχήματα. Η περιοχή στην οποία συμβαίνουν τα περισσότερα ατυχήματα είναι η “Br. Isles N. Sea E.Chnl Biscay”(24%) . Ακολουθεί η “China Japan & Korea”(20.7%) και η “South China & East Indies”(12.4%). Όλες οι υπόλοιπες περιοχές έχουν λιγότερα από 25 ατυχήματα(6.5%) ενώ οι περισσότερες παρουσιάζουν αμελητέο αριθμό ατυχημάτων.

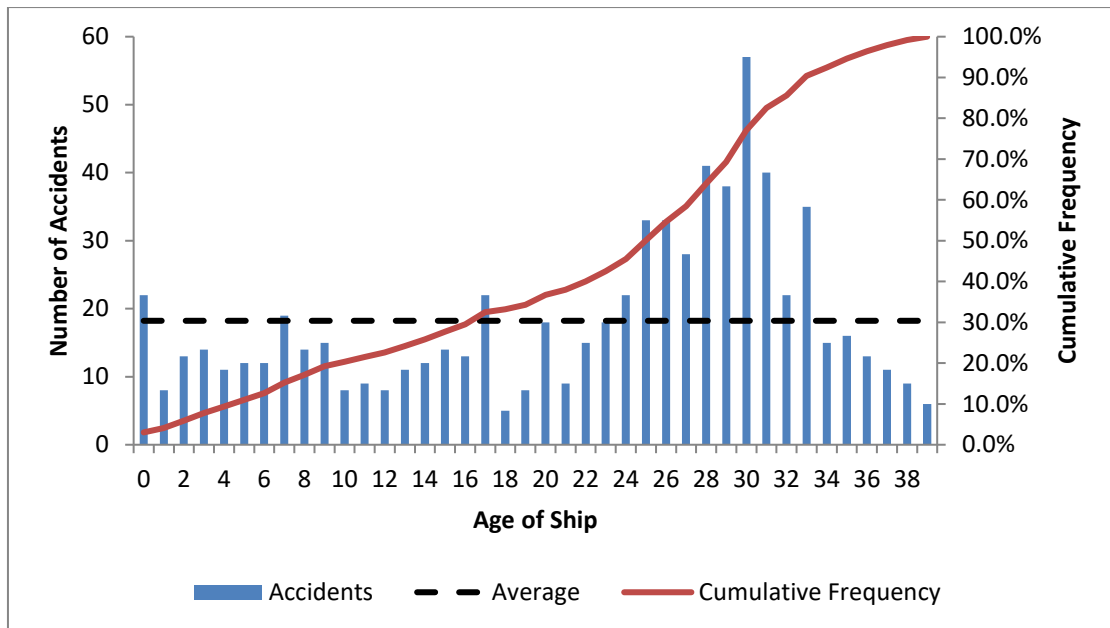
Στατιστική Ανάλυση Ναυτικών Ατυχημάτων σε πλοία LNG-LPG  
και Ανάπτυξη Αλγορίθμου Εξόρυξης Δεδομένων



Σχήμα 13 Συγκεντρωτική κατανομή ατυχημάτων συναρτήσει του Country Built

Στο Σχήμα 13 παρουσιάζεται η συγκεντρωτική κατανομή των ατυχημάτων συναρτήσει της χώρας κατασκευής του πλοίου. Ο μέσος όρος ατυχημάτων ανά χώρα κατασκευής είναι 25.7 ατυχήματα. Τα περισσότερα ατυχήματα συνολικά συνέβησαν σε πλοία κατασκευασμένα στην Ιαπωνία(42.4%) ενώ η χώρα κατασκευής με τα αμέσως επόμενα ατυχήματα είναι η Νότια Κορέα(24.9%). Όλες οι υπόλοιπες χώρες κατασκευής έχουν πολύ λίγα καταγεγραμμένα ατυχήματα στα πλοία τα οποία έχουν κατασκευάσει και σε καμία περίπτωση δεν υπερβαίνουν τα 20. Η Νότια Κορέα και η Κίνα εμφανίζουν μικρό ποσοστό ατυχημάτων, 24.9% και 5.1% σε σχέση με το ποσοστό του στόλου το οποίο έχουν κατασκευάσει 36.6% και 11.1% αντίστοιχα.

Στατιστική Ανάλυση Ναυτικών Ατυχημάτων σε πλοία LNG-LPG  
και Ανάπτυξη Αλγορίθμου Εξόρυξης Δεδομένων

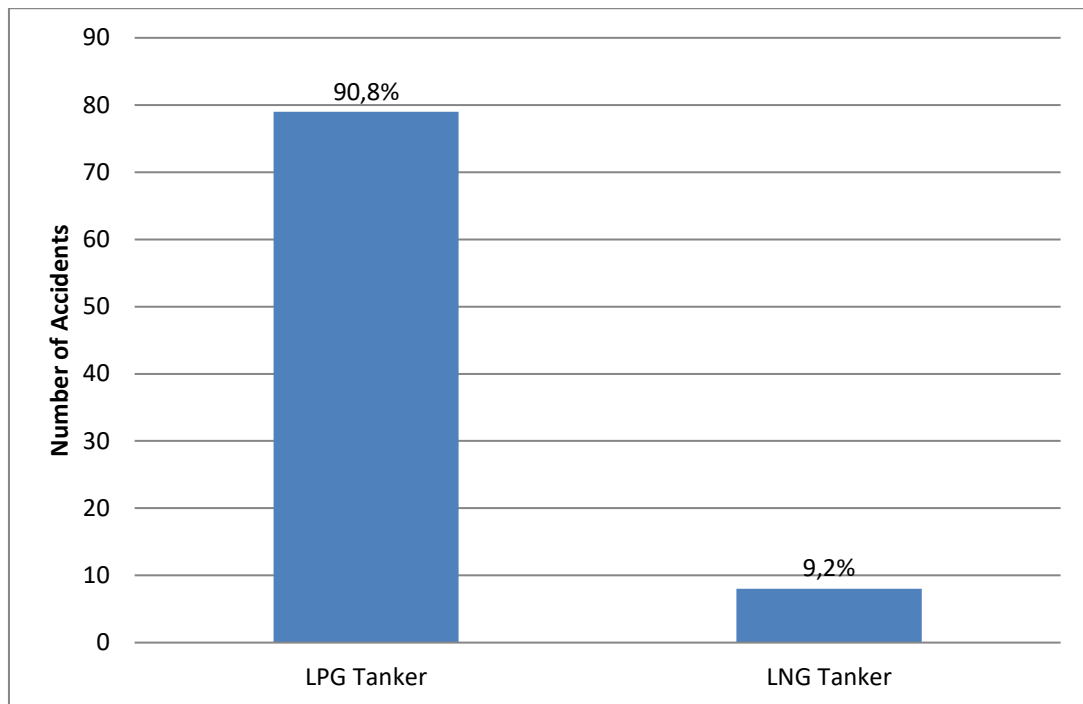


Σχήμα 14 Συγκεντρωτική κατανομή ατυχημάτων συναρτήσει του Age of Ship

Στο Σχήμα 14 παρουσιάζεται η συγκεντρωτική κατανομή των ατυχημάτων συναρτήσει της ηλικίας του πλοίου όταν συνέβη το ατύχημα. Ο μέσος όρος των ατυχημάτων ανά ηλικιακό έτος είναι 18.225 ατυχήματα. Περισσότερα ατυχήματα από τον μέσο όρο ανά ηλικιακό έτος εμφανίζονται κατά τον πρώτο χρόνο παραλαβής του πλοίου και επακόλουθα στο 7<sup>ο</sup>, 17<sup>ο</sup> και 24<sup>ο</sup> έως 33<sup>ο</sup> έτος ζωής του πλοίου. Τα περισσότερα ατυχήματα συμβαίνουν γύρω από το 30<sup>ο</sup> ηλικιακό έτος του πλοίου. Είναι ξεκάθαρο ότι όταν ένα υγραεριοφόρο πλοίο ξεπεράσει το 23<sup>ο</sup> έτος της ηλικίας του είναι πολύ πιο πιθανόν να υποστεί κάποιο ατύχημα. Λογικά αυτό οφείλεται στο γεγονός ότι τα πλοία σχεδιάζονται με ορίζοντα εικοσαετίας και μετά το πέρας αυτής της ηλικίας είναι αναπόφευκτο να αυξηθεί η ποσότητα των ατυχημάτων. Παράλληλα η πτώση που παρατηρείται στα ατυχήματα στην τελευταία πενταετία λειτουργίας των πλοίων οφείλεται στο γεγονός ότι η πλειονότητα του στόλου παροπλίζεται πριν φτάσει σε τέτοιες ηλικίες. Οπότε παρά το γεγονός ότι η πιθανότητα να συμβεί κάποιο ατύχημα βαίνει αύξουσα το Fleet at Risk βαίνει φθίνων. Τέλος ενδιαφέρον παρουσιάζει η μεγάλη συγκέντρωση ατυχημάτων κατά την παραλαβή του πλοίου στο 1<sup>ο</sup> έτος λειτουργίας του.

### Κεφάλαιο 4.1.3 Collision

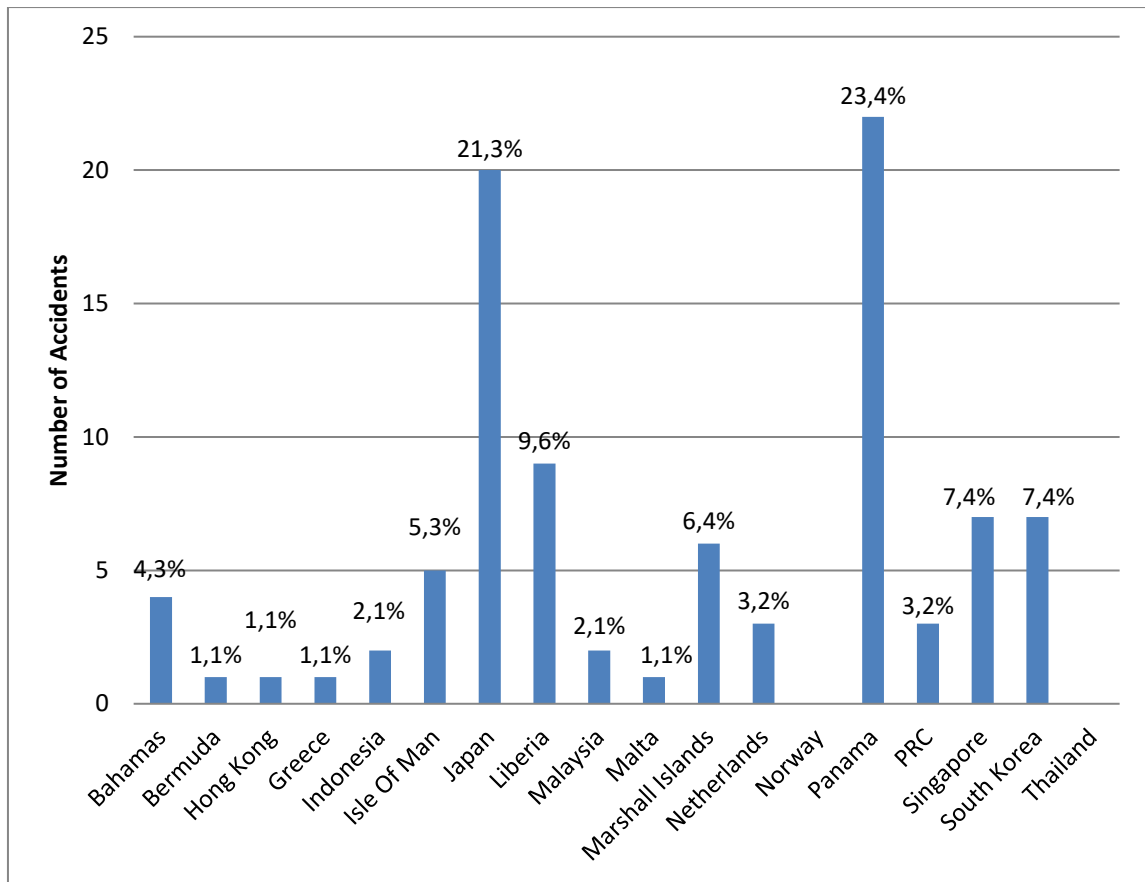
Σε αυτήν την κατηγορία συμπεριλαμβάνονται ατυχήματα κατά τα οποία υπάρχει σύγκρουση δύο πλοίων ανεξαρτήτως αν τα πλοία είναι εν πλω, αγκυροβολημένα ή δεμένα στην προβλήτα. Δεν περιλαμβάνονται ατυχήματα σε αυτήν την κατηγορία στα οποία υπήρξε πρόσκρουση ενός πλοίου με ένα ναυάγιο.[54]



Σχήμα 15 Κατανομή ατυχημάτων Collision συναρτήσει του Ship Type

Όπως φαίνεται στο Σχήμα 15 τα Collision ατυχήματα που έχουν συμβεί σε LPG Tankers(90.8%) είναι περίπου δεκαπλάσια σε σχέση με τα ατυχήματα που έχουν συμβεί στα LNG Tankers(9.2%). Αυτή η τεράστια διαφορά δεν συμβαδίζει με τα στοιχεία που εξήχθησαν για τον στόλο καθώς τα LPG Tankers αποτελούν το 73% του στόλου ενώ τα LNG Tankers το 27%. Είναι εμφανές ότι πλοία μεταφοράς LPG είναι πολύ πιο πιθανό να συναντήσουν κάποιο ατύχημα Collision από ότι πλοία μεταφοράς LNG.

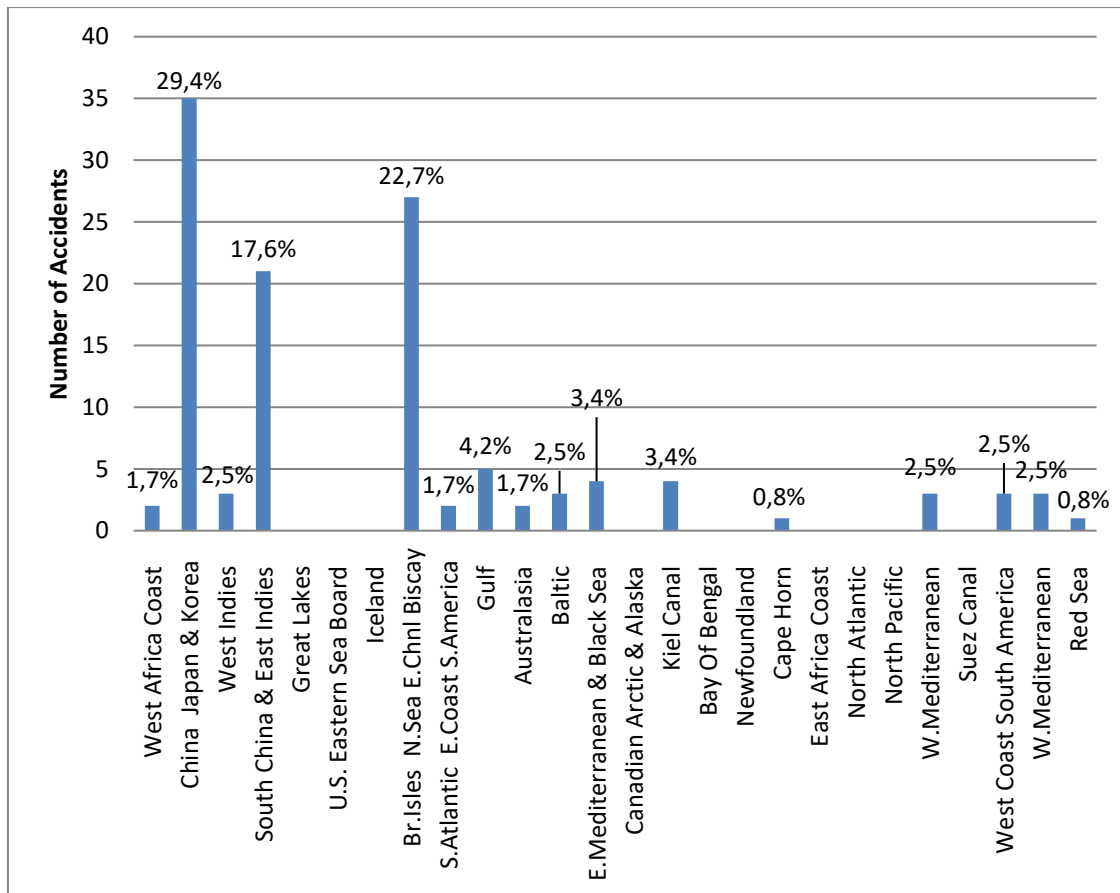
Στατιστική Ανάλυση Ναυτικών Ατυχημάτων σε πλοία LNG-LPG  
και Ανάπτυξη Αλγορίθμου Εξόρυξης Δεδομένων



Σχήμα 16 Κατανομή ατυχημάτων Collision συναρτήσει του Flag

Στο Σχήμα 16 παρατηρούμε ότι οι σημαίες που εμφανίζουν τα περισσότερα Collision ατυχήματα είναι η Ιαπωνία(21.3%) και ο Παναμάς(23.4%) με 20 και 22 ατυχήματα αντίστοιχα. Οι δύο αυτές χώρες έχουν υπό την σημαία τους το 8.8% και 16% του παγκόσμιου στόλου αντίστοιχα. Προφανώς το ποσοστό των ατυχημάτων Collision που καταλογίζεται στην σημαία της Ιαπωνίας είναι τεράστιο σε σχέση με τον στόλο.

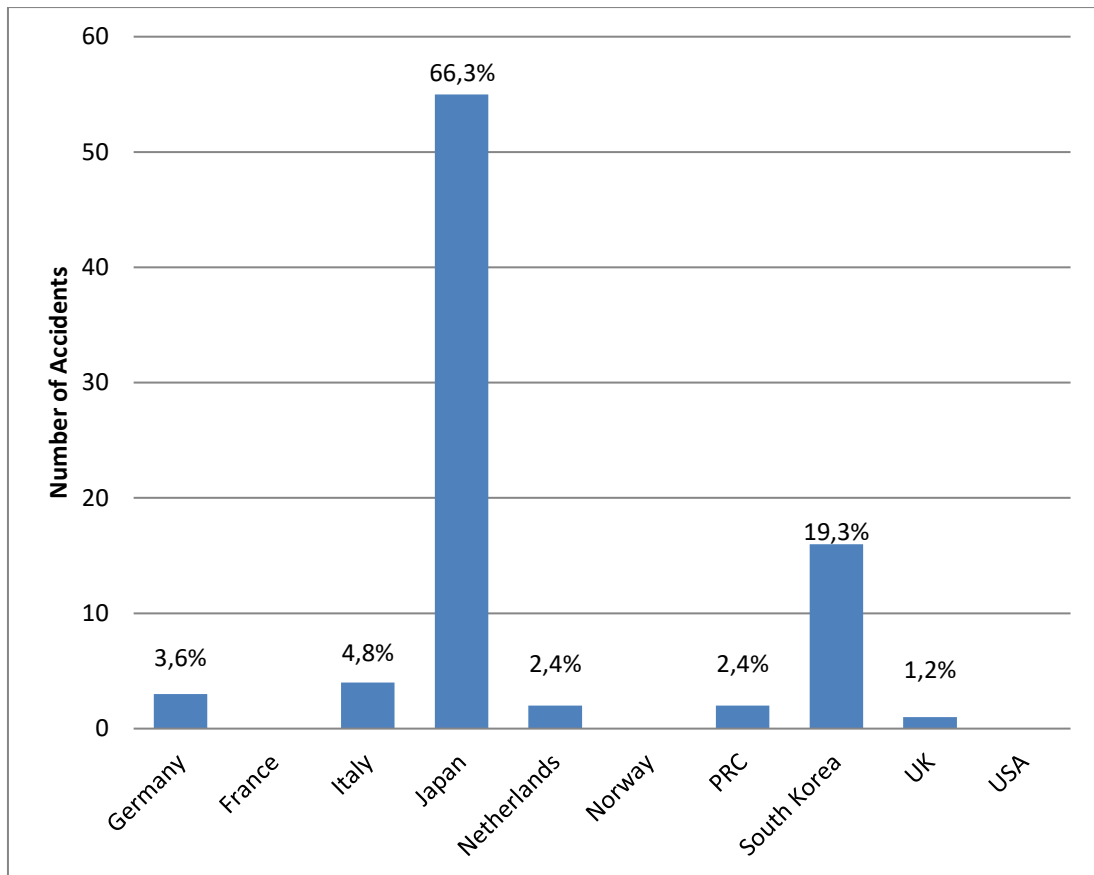
Στατιστική Ανάλυση Ναυτικών Ατυχημάτων σε πλοία LNG-LPG  
και Ανάπτυξη Αλγορίθμου Εξόρυξης Δεδομένων



Σχήμα 17 Κατανομή ατυχημάτων Collision συναρτήσει του Zone

Στο Σχήμα 17 εμφανίζεται η κατανομή των Collision ατυχημάτων συναρτήσει της ζώνης του ατυχήματος. Παρατηρούμε ότι τα περισσότερα Collision ατυχήματα έχουν συμβεί στις ακόλουθες περιοχές : China Japan & Korea(29.4%), Br. Isles N. Sea E.Chnl Biscay(22.7%), South China & East Indies(17.6%). Η κατανομή των ατυχημάτων Collision δεν ακολουθεί την συνολική κατανομή των ατυχημάτων για τις προαναφερθέντες περιοχές. Συγκεκριμένα τα ποσοστά που αντιστοιχούν σε αυτές είναι 20.7%, 24% και 12.4% αντίστοιχα. Στις υπόλοιπες περιοχές έχουν συμβεί πολύ λίγα ατυχήματα και δεν παρουσιάζεται κάποια πολύ σημαντική απόκλιση σε σχέση με την συνολική κατανομή των ατυχημάτων. Αξιοσημείωτο είναι το γεγονός ότι τα περισσότερα ατυχήματα έχουν συμβεί στην περιοχή “ Br. Isles N. Sea E.Chnl Biscay” ενώ τα περισσότερα Collision ατυχήματα έχουν συμβεί στην περιοχή “China Japan & Korea”.

Στατιστική Ανάλυση Ναυτικών Ατυχημάτων σε πλοία LNG-LPG  
και Ανάπτυξη Αλγορίθμου Εξόρυξης Δεδομένων

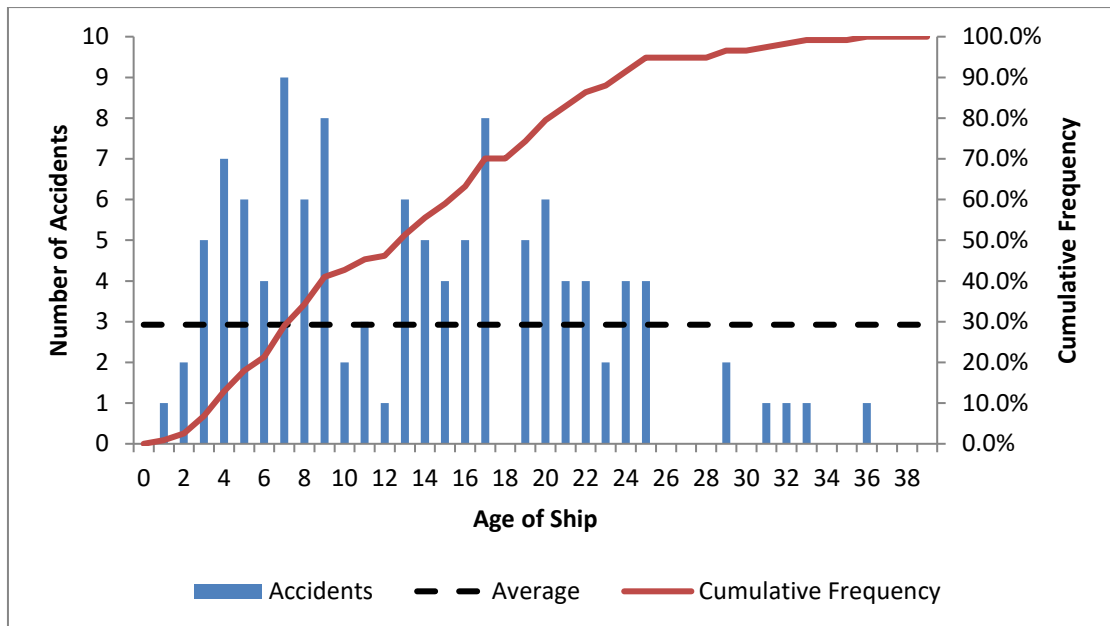


Σχήμα 18 Κατανομή ατυχημάτων Collision συναρτήσει του Country Built

Στο Σχήμα 18 εμφανίζεται η κατανομή των Collision ατυχημάτων συναρτήσει της χώρας κατασκευής του πλοίου. Περισσότερα από 50 πλοία (66.3%) που χτίστηκαν στην Ιαπωνία είχαν Collision ατύχημα ενώ μόλις το 44.8% του στόλου ναυπηγήθηκε στην Ιαπωνία. Αναλογικά με τα ατυχήματα των πλοίων που χτίστηκαν στην Νότιο Κορέα(36.6%) και στην Κίνα(11.1%) είναι ξεκάθαρο ότι πλοία χτισμένα στην Ιαπωνία έχουν την τάση να συμμετέχουν σε Collision ατυχήματα με μεγαλύτερη συχνότητα.



Στατιστική Ανάλυση Ναυτικών Ατυχημάτων σε πλοία LNG-LPG  
και Ανάπτυξη Αλγορίθμου Εξόρυξης Δεδομένων

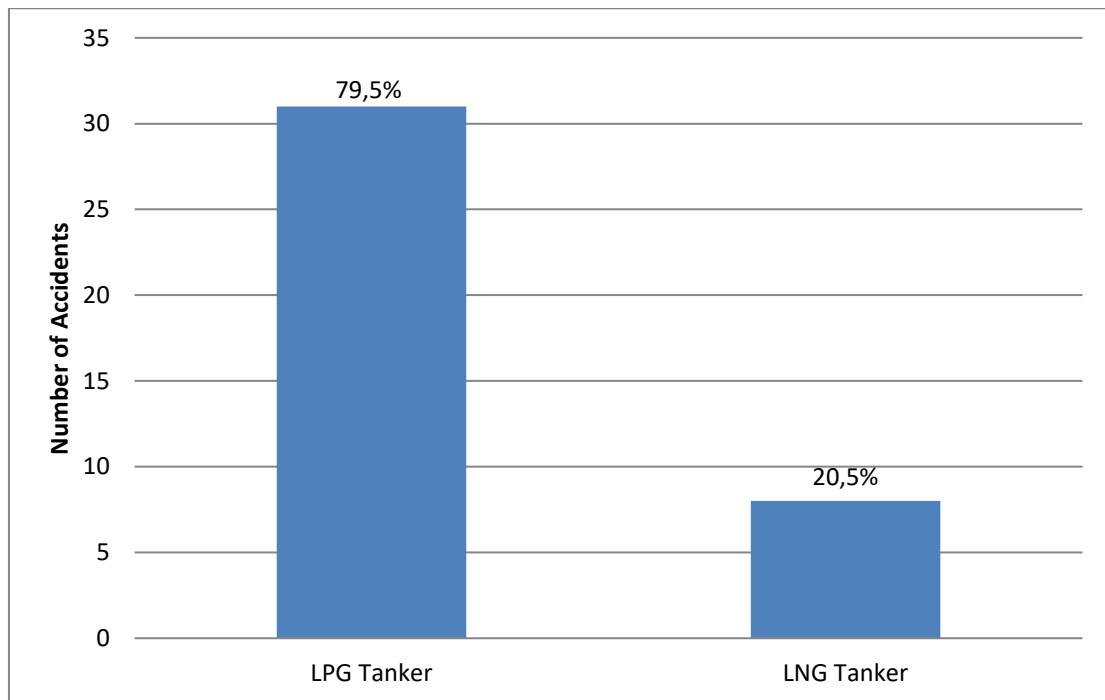


Σχήμα 19 Κατανομή ατυχημάτων Collision συναρτήσει του Age of Ship

Στο Σχήμα 19 παρουσιάζεται η κατανομή των ατυχημάτων Collision συναρτήσει της ηλικίας του πλοίου όταν συνέβη το ατύχημα. Ο μέσος όρος των ατυχημάτων ανά ηλικιακό έτος είναι 2.995 ατυχήματα. Παρατηρείται ότι η πλειονότητα των ατυχημάτων έχει συμβεί πριν το 26<sup>ο</sup> ηλικιακό έτος των πλοίων. Συγκεκριμένα τιμές ατυχημάτων ανά έτος που υπερβαίνουν την μέση τιμή ατυχημάτων παρατηρούνται από το 3<sup>ο</sup> έως 9<sup>ο</sup> ηλικιακό έτος και από το 13<sup>ο</sup> έως 25<sup>ο</sup> ηλικιακό έτος. Μετά το 25<sup>ο</sup> ηλικιακό έτος έχουν συμβεί ελάχιστα ατυχήματα Collision. Σε σχέση με την συγκεντρωτική κατανομή ατυχημάτων εμφανίζονται μεγάλες αποκλίσεις καθώς οι δύο αυτές κατανομές είναι εντελώς αντίθετες μεταξύ τους. Η γήρανση του πλοίου φαίνεται να μην αυξάνει την ύπαρξη ατυχημάτων Collision. Αντίθετα φαίνεται ότι η μείωση του Fleet at Risk με την πάροδο των ετών μειώνει τα συνολικά ατυχήματα Collision.

## Κεφάλαιο 4.1.4 Wrecked/Stranded

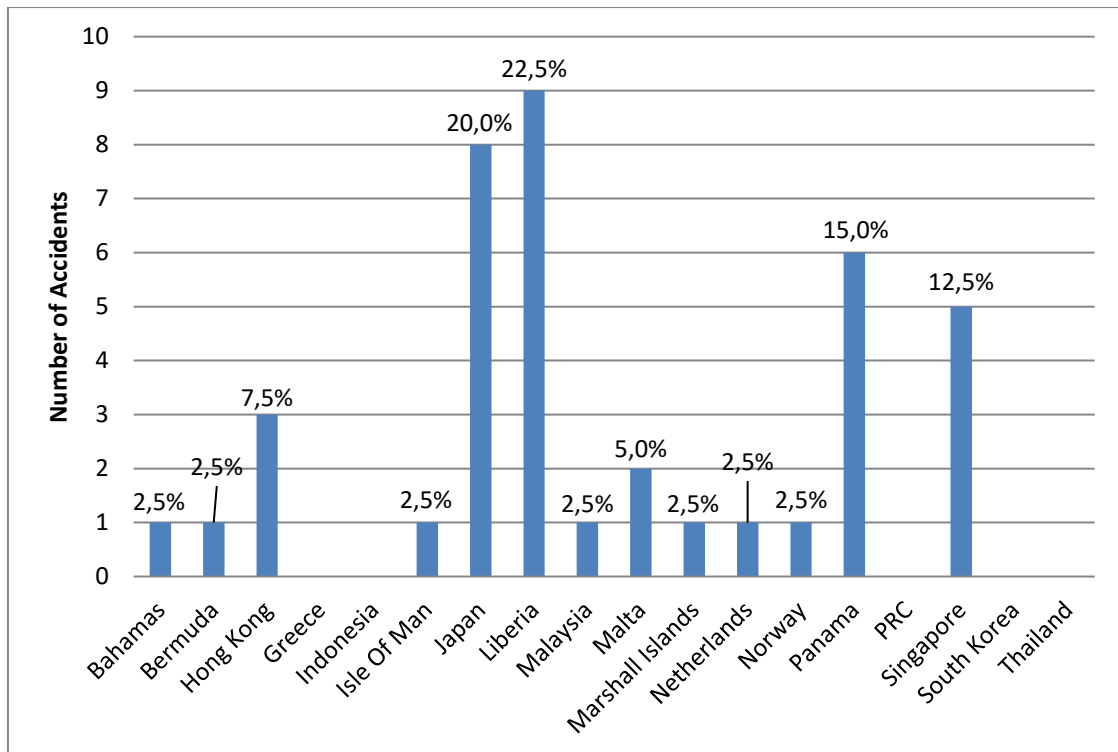
Σε αυτήν την κατηγορία συμπεριλαμβάνονται ατυχήματα στα οποία υπάρχει προσάραξη του πλοίου, πρόσκρουση του πλοίου με τον πυθμένα της θάλασσας (ύφαλοι) και πρόσκρουση του πλοίου με ναυάγια πλοίων.[54]



Σχήμα 20 Κατανομή ατυχημάτων Wrecked/Stranded συναρτήσει του Ship Type

Στο Σχήμα 20 εμφανίζεται η κατανομή των Wrecked/Stranded ατυχημάτων συναρτήσει του είδους του πλοίου. Παρατηρούμε ότι τα ατυχήματα που συνέβησαν σε LPG Tankers(79.5%) είναι τετραπλάσια από τα ατυχήματα που συνέβησαν σε LNG Tankers(20.5%). Στο σύνολο του στόλου που μελετάται το 73% του στόλου αποτελείται από LPG Tankers ενώ το υπόλοιπο είναι LNG. Οι διαφορές που εμφανίζονται μεταξύ στόλου και ατυχημάτων δεν είναι ιδιαίτερα μεγάλες.

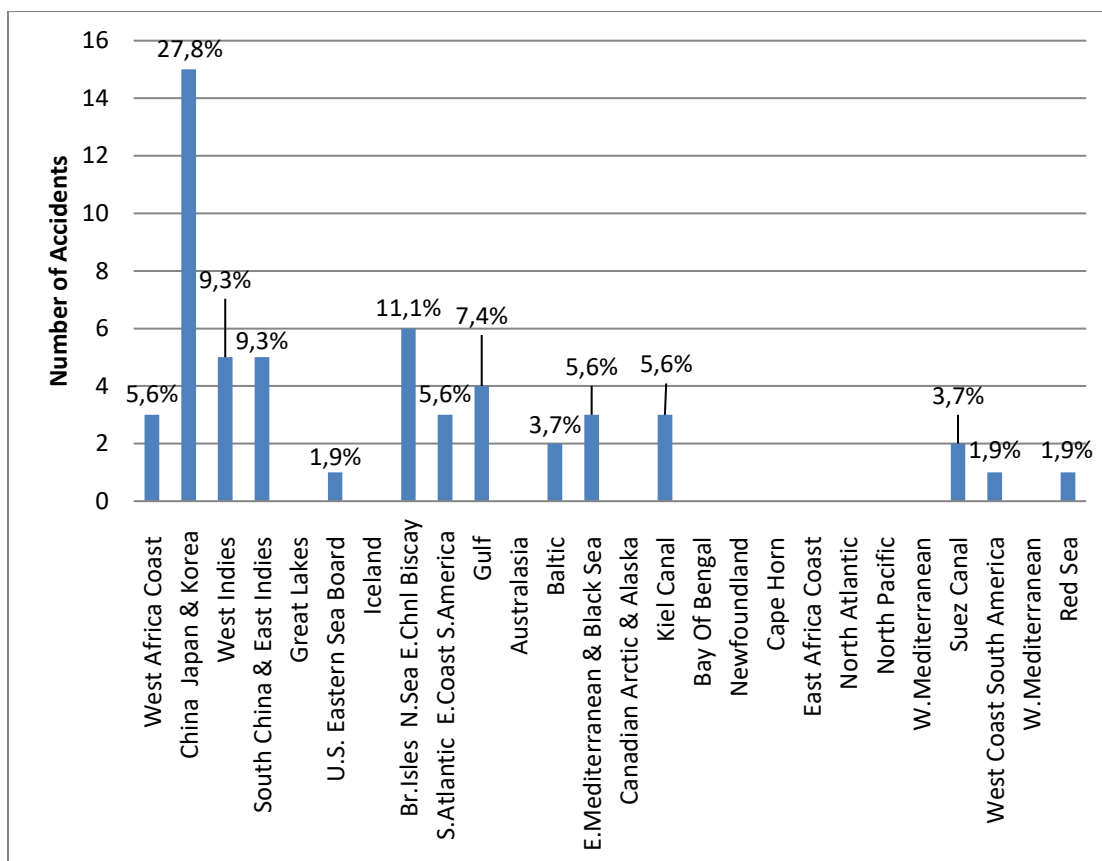
Στατιστική Ανάλυση Ναυτικών Ατυχημάτων σε πλοία LNG-LPG  
και Ανάπτυξη Αλγορίθμου Εξόρυξης Δεδομένων



Σχήμα 21 Κατανομή ατυχημάτων Wrecked/Stranded συναρτήσει του Flag

Στο Σχήμα 21 εμφανίζεται η κατανομή των Wrecked/Stranded ατυχημάτων συναρτήσει της σημαίας του πλοίου. Σημαντικός αριθμός ατυχημάτων Wrecked/Stranded συνέβη σε πλοία με σημαίες: Ιαπωνία(20%), Λιβερία(22.5%), Παναμά(15%) και Σιγκαπούρη(12.5%). Οι μεγαλύτερες αποκλίσεις σε σχέση με την κατανομή του στόλου εμφανίζονται στην σημαία της Λιβερίας στην οποία υπάγεται το 8.1% του στόλου, της Ιαπωνίας που έχει το 8.8% του στόλου, των νήσων Μάρσαλ με 9.2% του στόλου και της Σιγκαπούρης με 8.2% του στόλου.

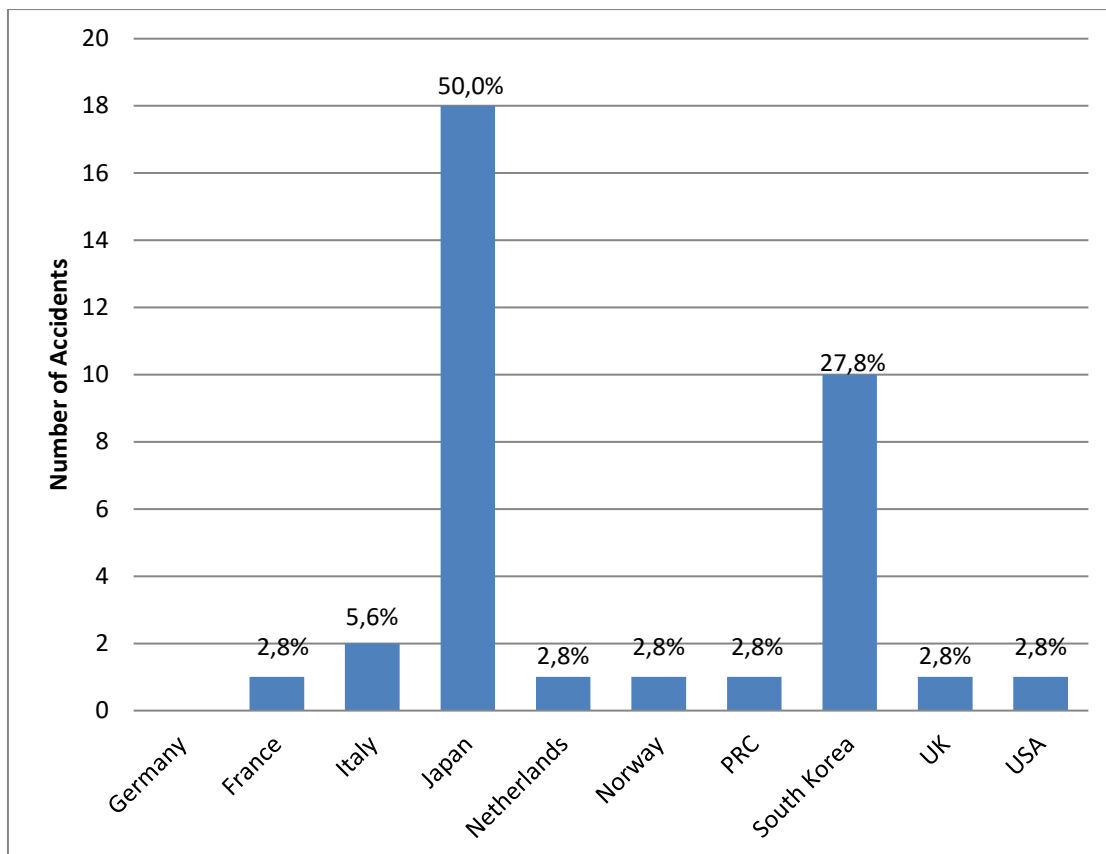
Στατιστική Ανάλυση Ναυτικών Ατυχημάτων σε πλοία LNG-LPG  
και Ανάπτυξη Αλγορίθμου Εξόρυξης Δεδομένων



Σχήμα 22 Κατανομή ατυχημάτων Wrecked/Stranded συναρτήσει του Zone

Στο Σχήμα 22 εμφανίζεται η κατανομή των Wrecked/Stranded ατυχημάτων συναρτήσει της ζώνης του ατυχήματος. Παρατηρούμε στο διάγραμμα ότι η περιοχή που εμφανίζει τα περισσότερα Wrecked/Stranded ατυχήματα είναι “China Japan & Korea”(27.8%). Τα αμέσως περισσότερα ατυχήματα έχουν συμβεί στις περιοχές “Br. Isles N. Sea E.Chnl Biscay”(11.7%), “South China & East Indies”(9.3%) και “West Indies”(9.3%). Σε σχέση με την συνολική κατανομή των ατυχημάτων η περιοχή “Br. Isles N. Sea E.Chnl Biscay” εμφανίζει λιγότερα από τα αναμενόμενα ατυχήματα καθώς σε αυτήν την περιοχή έχει συμβεί το 24% των συνολικών ατυχημάτων. Αντίθετα στις περιοχές “West Indies” και “China Japan & Korea” στις οποίες έχουν συμβεί το 4.3% και 20.7% των συνολικών ατυχημάτων αντίστοιχα τα ατυχήματα Wrecked/Stranded είναι περισσότερα από τα αναμενόμενα.

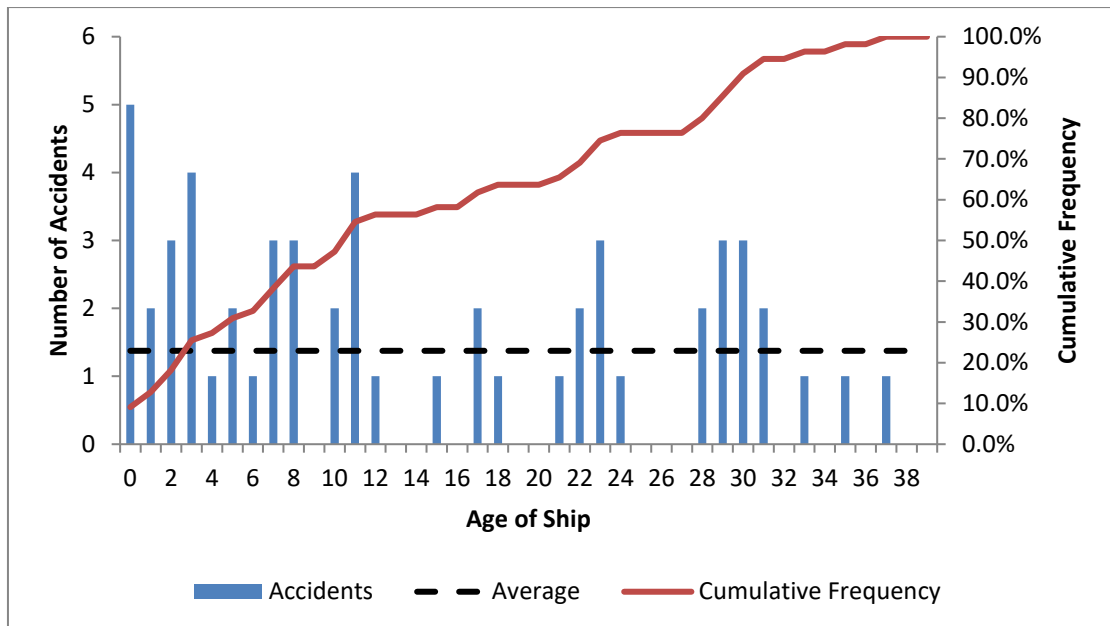
Στατιστική Ανάλυση Ναυτικών Ατυχημάτων σε πλοία LNG-LPG  
και Ανάπτυξη Αλγορίθμου Εξόρυξης Δεδομένων



Σχήμα 23 Κατανομή ατυχημάτων Wrecked/Stranded συναρτήσεϊ του Country Built

Στο Σχήμα 23 εμφανίζεται η κατανομή των Wrecked/Stranded ατυχημάτων συναρτήσεϊ της χώρας κατασκευής του πλοίου. Η πλειονότητα των Wrecked/Stranded ατυχημάτων συνέβη σε πλοία τα οποία κατασκευάστηκαν στην Ιαπωνία(50%) και στην Νότια Κορέα(27.8%). Σε σύγκριση με την κατανομή του στόλου η Ιαπωνία έχει περισσότερα από τα αναμενόμενα ατυχήματα καθώς έχει ναυπηγήσει το 44.8% του στόλου ενώ η Νότια Κορέα και η Κίνα έχουν λιγότερα ατυχήματα από τα αναμενόμενα έχοντας ναυπηγήσει το 36.6% και το 11.1% του στόλου αντίστοιχα.

Στατιστική Ανάλυση Ναυτικών Ατυχημάτων σε πλοία LNG-LPG  
και Ανάπτυξη Αλγορίθμου Εξόρυξης Δεδομένων

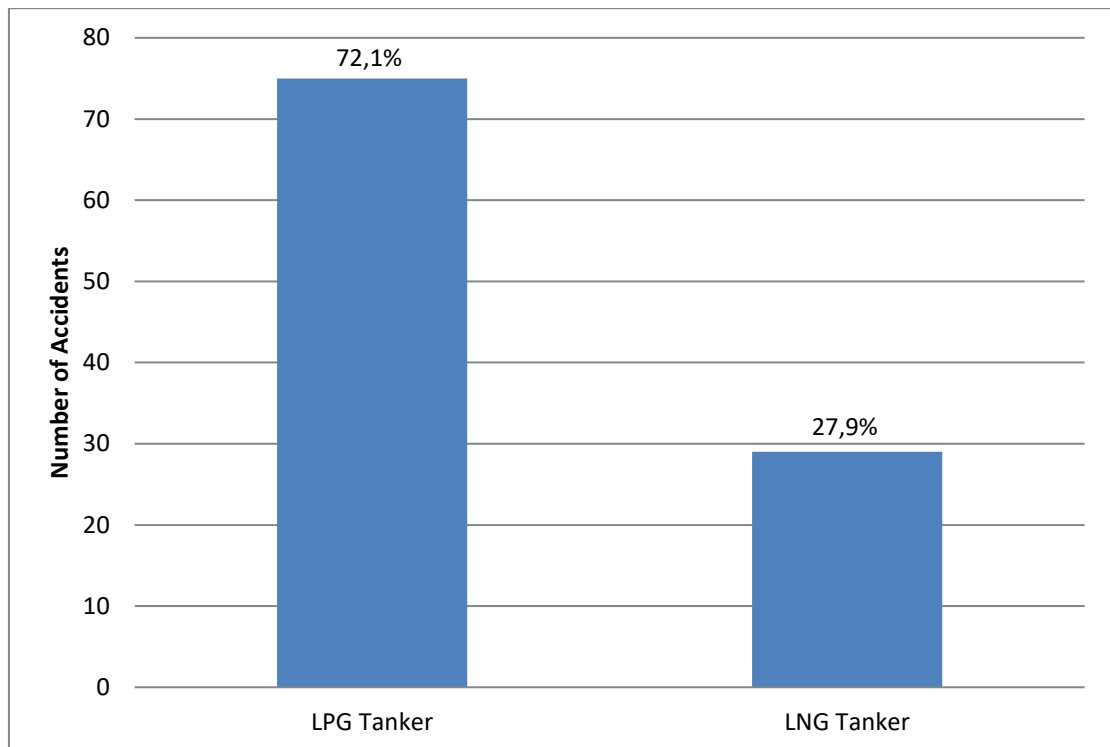


Σχήμα 24 Κατανομή ατυχημάτων Wrecked/Stranded συναρτήσει του Age of Ship

Στο Σχήμα 24 παρουσιάζεται η κατανομή των ατυχημάτων Wrecked/Stranded συναρτήσει της ηλικίας του πλοίου όταν συνέβη το ατύχημα. Ο μέσος όρος των ατυχημάτων ανά ηλικιακό έτος είναι 1.375 ατυχήματα. Η μεγαλύτερη συγκέντρωση ατυχημάτων Wrecked/Stranded που ξεπερνούν τον μέσο όρο εμφανίζεται στην πρώτη δεκαετία λειτουργίας του πλοίου και συγκεκριμένα το μέγιστο στον 1<sup>ο</sup> χρόνο λειτουργίας του. Μετά την πάροδο των 10 πρώτων ετών εμφανίζονται ορισμένα ηλικιακά έτη ( 17<sup>ο</sup>, 22<sup>ο</sup>, 23<sup>ο</sup>, και 28<sup>ο</sup> έως 31<sup>ο</sup>) όπου τα ατυχήματα ξεπερνούν τον μέσο όρο. Επειδή τα συνολικά ατυχήματα τα οποία έχουν συμβεί σε αυτήν την κατηγορία είναι σχετικά λίγα αυτές οι σποραδικές αυξήσεις των ατυχημάτων μπορούν να αποδοθούν στην τυχαιότητα και δεν εκφράζουν κάποια συγκεκριμένη τάση. Η κύρια συγκέντρωση ατυχημάτων εμφανίζεται ξεκάθαρα στην πρώτη δεκαετία λειτουργίας του πλοίου.

## Κεφάλαιο 4.1.5 Hull & Machinery (Mchy) Damage

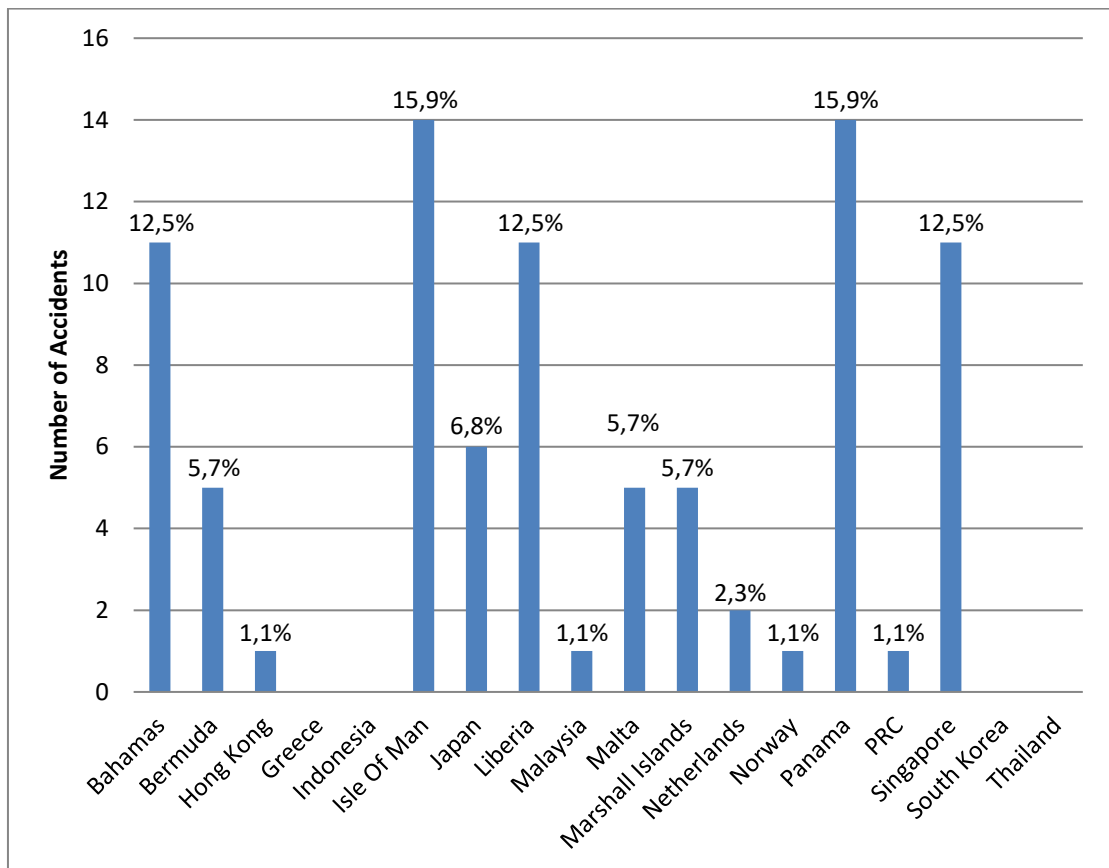
Σε αυτήν την κατηγορία συμπεριλαμβάνονται ατυχήματα στα οποία υπήρξε βλάβη στην γάστρα ή και την μηχανολογική εγκατάσταση του πλοίου. Εξαιρούνται ατυχήματα στα οποία η εν λόγω βλάβη προέκυψε λόγω σύγκρουσης, προσάραξης ή οποιασδήποτε άλλης κατηγορίας ατυχήματος.[54]



Σχήμα 25 Κατανομή ατυχημάτων Hull&Mchy Damage συναρτήσει του Ship Type

Στο Σχήμα 25 εμφανίζεται η κατανομή των Hull&Mchy. Damage ατυχημάτων συναρτήσει του τύπου του πλοίου. Τα ατυχήματα τα οποία συνέβησαν σε LNG Tankers αποτελούν το 27.9% ενώ τα ατυχήματα που συνέβησαν σε LPG Tankers το 72.1%. Συνολικά παρατηρείται μεγάλη σύγκλιση με τις τιμές που έχουμε για τον στόλο 27% και 73% αντίστοιχα και συνεπώς η ύπαρξη ή μη Hull&Mchy. Damage ατυχημάτων φαίνεται να είναι ανεξάρτητη από τον τύπο του πλοίου.

Στατιστική Ανάλυση Ναυτικών Ατυχημάτων σε πλοία LNG-LPG  
και Ανάπτυξη Αλγορίθμου Εξόρυξης Δεδομένων

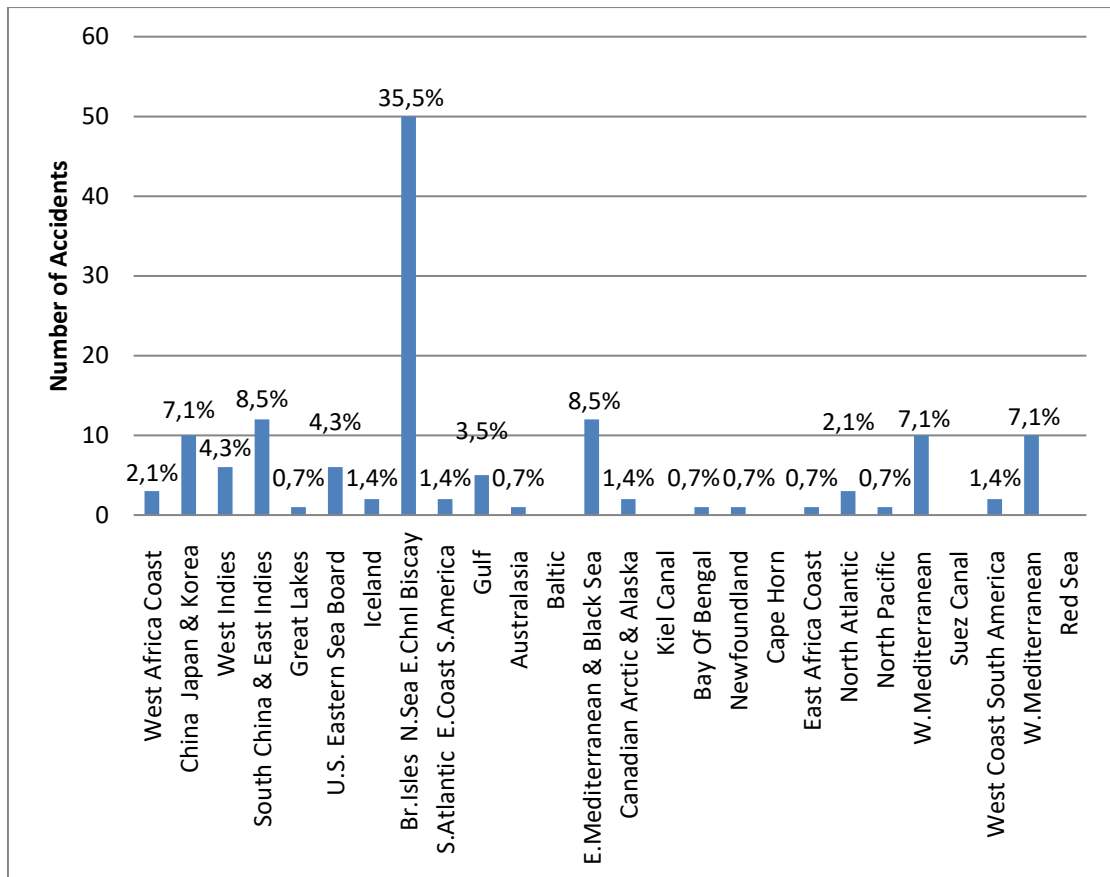


Σχήμα 26 Κατανομή ατυχημάτων Hull&Mchy Damage συναρτήσεσι του Flag

Στο Σχήμα 26 εμφανίζεται η κατανομή των Hull&Mchy. Damage ατυχημάτων συναρτήσεσι της σημαίας του πλοίου. Παρατηρούμε ότι τα πλοία με τις ακόλουθες σημαίες ενεπλάκησαν στα περισσότερα Hull&Mchy. Damage ατυχήματα: Isle of Man(15.9%), Παναμάς(15.9%), Λιβερία(12.5%), Σιγκαπούρη και Μπαχάμες(12.5%). Οι μεγαλύτερες αυξήσεις ατυχημάτων συγκριτικά με την κατανομή του στόλου εμφανίζονται στις σημαίες του Isle of Man που έχει 4% του στόλου, των Μπαχαμών με 7.1% του στόλου, της Λιβερίας με 8.1% του στόλου και της Σιγκαπούρης με 8.2% του στόλου. Αντίθετα οι σημαίες της Ιαπωνίας και των νήσων Μάρσαλ εμφανίζουν λιγότερα ατυχήματα από τα αναμενόμενα καθώς έχουν το 8.8% και 9.2% του στόλου αντίστοιχα και μόλις το 6.8% και 5.7% των ατυχημάτων.



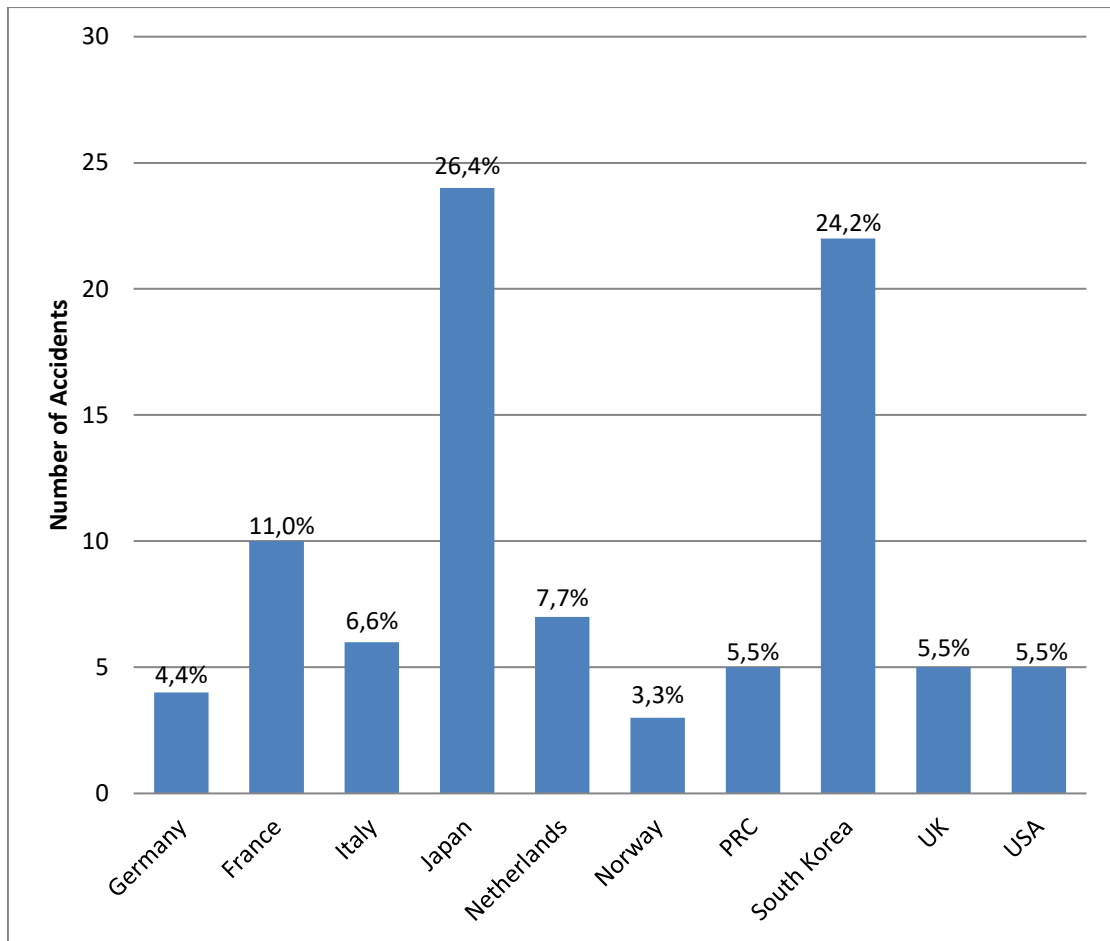
Στατιστική Ανάλυση Ναυτικών Ατυχημάτων σε πλοία LNG-LPG  
και Ανάπτυξη Αλγορίθμου Εξόρυξης Δεδομένων



Σχήμα 27 Κατανομή ατυχημάτων Hull&Mchy Damage συναρτήσει του Zone

Στο Σχήμα 27 εμφανίζεται η κατανομή των Hull&Mchy. Damage ατυχημάτων συναρτήσει της ζώνης του ατυχήματος. Παρατηρούμε ότι η μοναδική περιοχή η οποία εμφανίζει υψηλές τιμές ύπαρξης ατυχημάτων Hull&Mchy. Damage είναι η “Br. Isles N. Sea E.Chnl Biscay” με 50 συνολικά ατυχήματα(35.5%). Συγκριτικά με την συνολική κατανομή των ατυχημάτων παρατηρείται ότι η περιοχή “Br. Isles N. Sea E.Chnl Biscay” παρουσιάζει μεγάλο ποσοστό σε σχέση με τα συνολικά(24.0%) ενώ οι περιοχές “China Japan & Korea” και “South China & East Indies” παρουσιάζουν πολύ λίγα ατυχήματα Hull&Mchy Damage σε σχέση με τα συνολικά τους νούμερα, 20.7% και 12.4% αντίστοιχα.

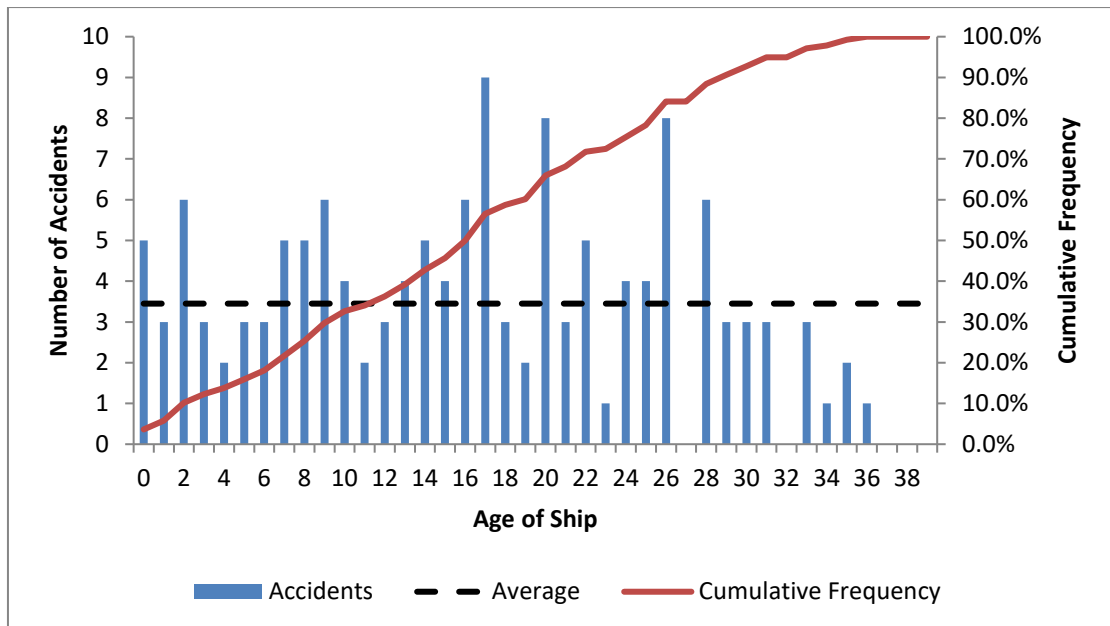
Στατιστική Ανάλυση Ναυτικών Ατυχημάτων σε πλοία LNG-LPG  
και Ανάπτυξη Αλγορίθμου Εξόρυξης Δεδομένων



Σχήμα 28 Κατανομή ατυχημάτων Hull&Mch. Damage συναρτήσεϊ του Country Built

Στο Σχήμα 28 εμφανίζεται η κατανομή των Hull&Mch. Damage ατυχημάτων συναρτήσεϊ της χώρας κατασκευής του πλοίου. Στο διάγραμμα παρατηρούμε ότι οι χώρες κατασκευής με την μεγαλύτερη συγκέντρωση ατυχημάτων είναι η Νότια Κορέα(24.2%) και η Ιαπωνία(26.4%). Συγκρινόμενα αυτά τα μεγέθη με την κατανομή του στόλου παρατηρούμε ότι οι 3 κύριες χώρες κατασκευής Ιαπωνία(44.8%), Νότια Κορέα(36.6%) και Κίνα(11.1%) εμφανίζουν πολύ λίγα Hull&Mch. Damage ατυχήματα αναλογικά ενώ οι υπόλοιπες χώρες που έχουν κατασκευάσει μικρό ποσοστό του στόλου έχουν αυξημένα ποσοστά.

Στατιστική Ανάλυση Ναυτικών Ατυχημάτων σε πλοία LNG-LPG  
και Ανάπτυξη Αλγορίθμου Εξόρυξης Δεδομένων

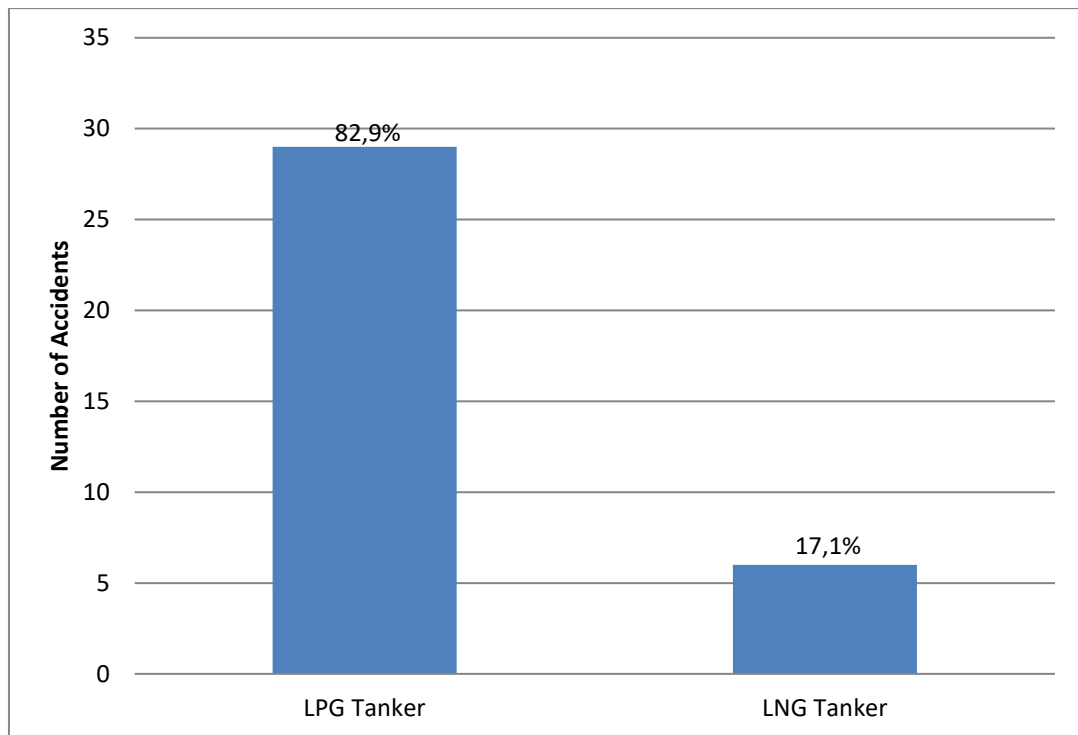


Σχήμα 29 Κατανομή ατυχημάτων Hull&MchY Damage συναρτήσει του Age of Ship

Στο Σχήμα 29 παρουσιάζεται η κατανομή των ατυχημάτων Hull&MchY Damage συναρτήσει της ηλικίας του πλοίου όταν συνέβη το ατύχημα. Ο μέσος όρος των ατυχημάτων ανά ηλικιακό έτος είναι 3.45 ατυχήματα. Οι υψηλότερες τιμές ατυχημάτων εμφανίζονται στο 16<sup>ο</sup>, 20<sup>ο</sup> και 26<sup>ο</sup> ηλικιακό έτος των πλοίων. Συνολικά δεν παρατηρείται κάποια ηλικιακή περίοδος όπου τα ατυχήματα ανά έτος ξεπερνάνε σταθερά την μέση τιμή ατυχημάτων. Αντίθετα φαίνεται ότι τα ατυχήματα ανά έτος παραμένουν σταθερά γύρω από την μέση τιμή με μικρές αποκλίσεις από την κατασκευή έως το 33<sup>ο</sup> ηλικιακό έτος των πλοίων. Σε σχέση με την συνολική κατανομή των ατυχημάτων συναρτήσει της ηλικίας του πλοίου τα αποτελέσματα βρίσκονται σε κάποια αντιδιαστολή καθώς δεν παρατηρείται ιδιαίτερη αύξηση των ατυχημάτων Hull&MchY Damage γύρω από το 30<sup>ο</sup> έτος ηλικίας των πλοίων.

## Κεφάλαιο 4.1.6 Fire/Explosion

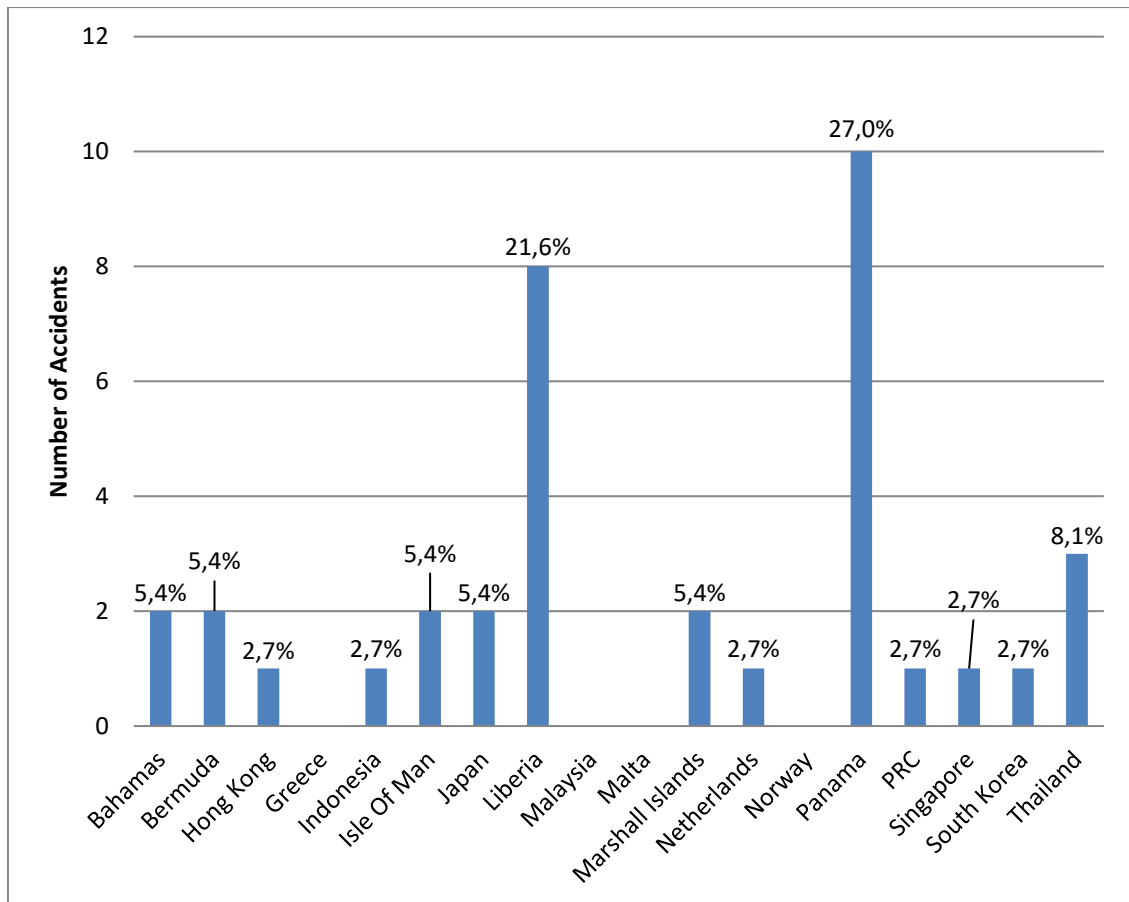
Σε αυτήν την κατηγορία συμπεριλαμβάνονται ατυχήματα στα οποία υπήρξε πυρκαγιά ή έκρηξη λόγω μηχανολογικής βλάβης ή ατυχήματα στα οποία η πυρκαγιά / έκρηξη ήταν το πρώτο συμβάν που αναφέρθηκε. Ατυχήματα στα οποία η πυρκαγιά ή η έκρηξη οφείλεται σε σύγκρουση, προσάραξη κλπ. δεν συμπεριλαμβάνονται σε αυτήν την κατηγορία.[54]



Σχήμα 30 Κατανομή ατυχημάτων Fire&Explosion συναρτήσει του Ship Type

Στο Σχήμα 30 εμφανίζεται η κατανομή των Fire&Explosion ατυχημάτων συναρτήσει του τύπου του πλοίου. Στα LPG Tankers έχουν συνολικά συμβεί 28 ατυχήματα (82.9%) ενώ στα LNG Tankers έχουν συνολικά συμβεί 6 ατυχήματα (17.1%). Στατιστικά η κατανομή των ατυχημάτων δεν ακολουθεί την κατανομή του στόλου καθώς από τα καταγεγραμμένα πλοία το 73% είναι LPG Tankers ενώ το 27% είναι LNG Tankers. Είναι πιο πιθανόν συνεπώς να συμβεί κάποια έκρηξη η πυρκαγιά σε πλοίο μεταφοράς LPG.

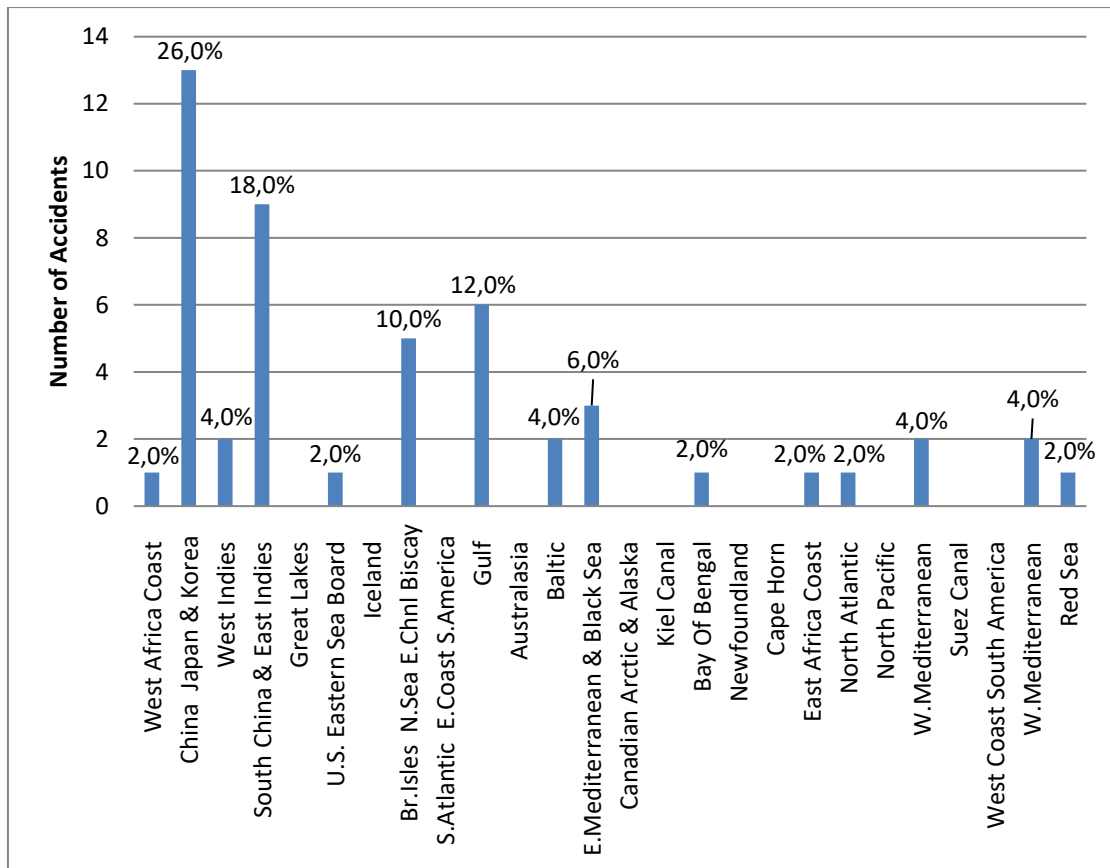
Στατιστική Ανάλυση Ναυτικών Ατυχημάτων σε πλοία LNG-LPG  
και Ανάπτυξη Αλγορίθμου Εξόρυξης Δεδομένων



Σχήμα 31 Κατανομή ατυχημάτων Fire&Explosion συναρτήσεως του Flag

Στο Σχήμα 31 εμφανίζεται η κατανομή των Fire&Explosion ατυχημάτων συναρτήσεως της σημαίας του πλοίου. Στο ανωτέρω διάγραμμα φαίνεται ότι η πλειονότητα των Fire & Explosion ατυχημάτων συνέβησαν σε πλοία τα οποία έφεραν σημαία του Παναμά (27%) ή της Λιβερίας (21.6%). Και οι δύο αυτές σημαίες συγκεντρώνουν μεγάλο αριθμό ατυχημάτων αναλογικά με τα πλοία τα οποία έχουν, 16% και 9.2% αντίστοιχα. Ταυτόχρονα οι σημαίες της Ιαπωνίας, Σιγκαπούρης, των νήσων Μαρσαλ και των Μπαχαμών με ποσοστά στόλου 8.8%, 8.2%, 9.2%, 7.1% αντίστοιχα συγκεντρώνουν χαμηλότερα ποσοστά ατυχημάτων από τα αναμενόμενα όπως φαίνεται στο Σχήμα 31.

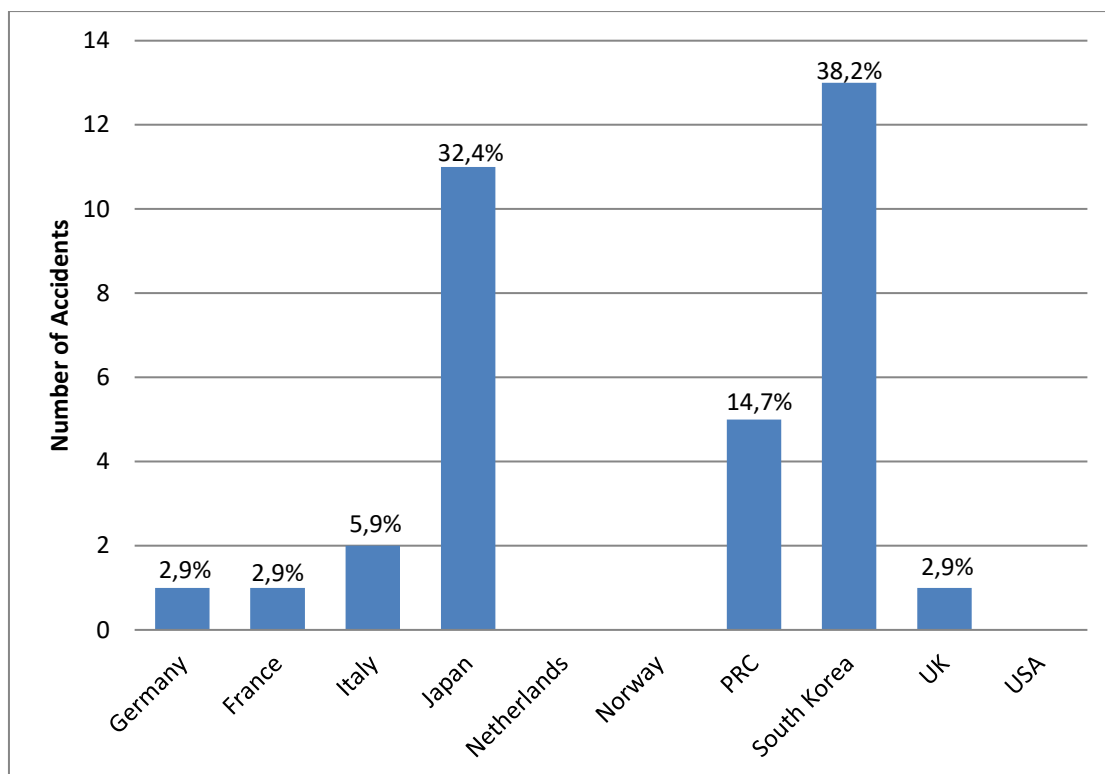
Στατιστική Ανάλυση Ναυτικών Ατυχημάτων σε πλοία LNG-LPG  
και Ανάπτυξη Αλγορίθμου Εξόρυξης Δεδομένων



Σχήμα 32 Κατανομή ατυχημάτων Fire&Explosion συναρτήσει του Zone

Στο Σχήμα 32 εμφανίζεται η κατανομή των Fire&Explosion ατυχημάτων συναρτήσει της ζώνης του ατυχήματος. Οι περιοχές οι οποίες εμφανίζουν τα περισσότερα Fire&Explosion ατυχήματα είναι οι: “South China & East Indies”(18%) και “China Japan & Korea”(26%) ενώ ακολουθούν οι περιοχές του “Κόλπου του Μεξικού”(12%) και “Br. Isles N. Sea E. Chnl Biscay”(10%). Οι περιοχές “South China & East Indies”, “China Japan & Korea” και “Κόλπος του Μεξικού” με κατανομές συνολικών ατυχημάτων ίσες με 12.4%, 20.7% και 5.3% αντίστοιχα εμφανίζουν περισσότερα από τα προβλεπόμενα ατυχήματα ενώ η περιοχή “Br. Isles N. Sea E. Chnl Biscay” με συνολικά ατυχήματα της τάξης του 24% εμφανίζει πολύ λιγότερα.

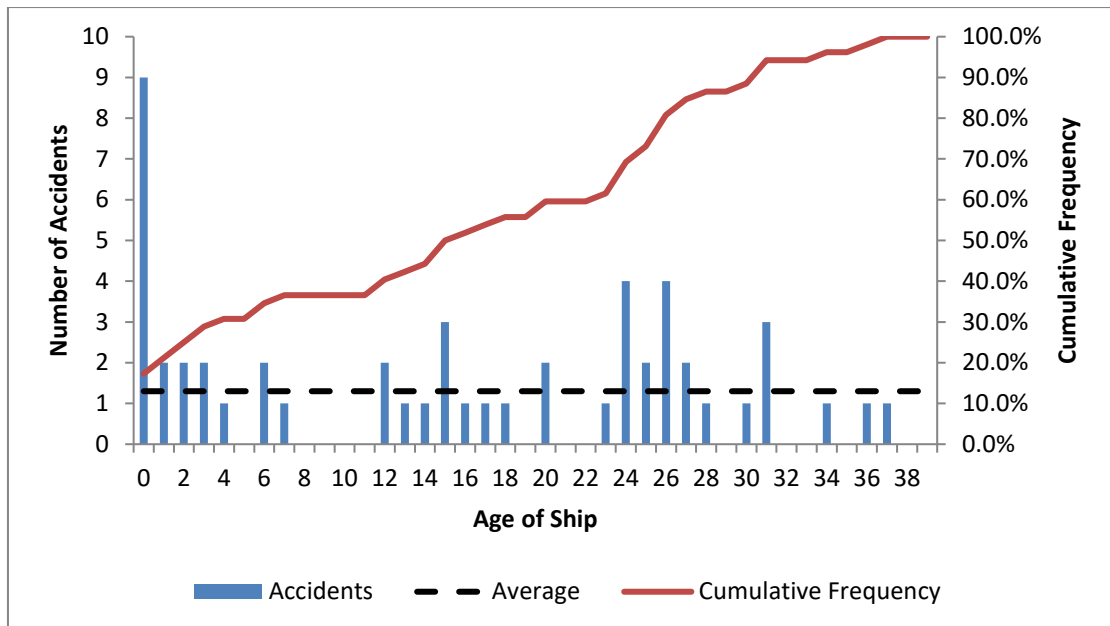
Στατιστική Ανάλυση Ναυτικών Ατυχημάτων σε πλοία LNG-LPG  
και Ανάπτυξη Αλγορίθμου Εξόρυξης Δεδομένων



Σχήμα 33 Κατανομή ατυχημάτων Fire&Explosion συναρτήσει του Country Built

Στο Σχήμα 33 εμφανίζεται η κατανομή των Fire&Explosion ατυχημάτων συναρτήσει της χώρας κατασκευής του πλοίου. Οι χώρες που εμφανίζουν τα μεγαλύτερα ποσοστά ατυχημάτων είναι η Νότιος Κορέα, η Ιαπωνία και η Κίνα με ποσοστά 38.2%, 32.4% και 14.7% αντίστοιχα. Τα ποσοστά του στόλου τα οποία έχουν κατασκευαστεί σε αυτές τις τρεις χώρες είναι 36.6%, 44.8% και 11.1% αντίστοιχα. Παρατηρείται λοιπόν ότι Νότιος Κορέα και η Κίνα έχουν ελαφρώς περισσότερα ατυχήματα από τα αναμενόμενα ενώ η Ιαπωνία έχει πολύ λιγότερα από τα ατυχήματα που τις αναλογούν.

Στατιστική Ανάλυση Ναυτικών Ατυχημάτων σε πλοία LNG-LPG  
και Ανάπτυξη Αλγορίθμου Εξόρυξης Δεδομένων



Σχήμα 34 Κατανομή ατυχημάτων Fire&Explosion συναρτήσει του Age of Ship

Στο Σχήμα 34 παρουσιάζεται η κατανομή των ατυχημάτων Fire&Explosion συναρτήσει της ηλικίας του πλοίου όταν συνέβη το ατύχημα. Ο μέσος όρος των ατυχημάτων ανά ηλικιακό έτος είναι 1.3 ατυχήματα. Τα περισσότερα ατυχήματα Fire & Explosion εμφανίζονται στον πρώτο χρόνο λειτουργίας του πλοίου. Το γεγονός αυτό οφείλεται κατά πάσα πιθανότητα στην άγνοια του πληρώματος και της διαχειριστικής εταιρίας για το πλοίο που οδηγεί σε προβλήματα και εν τέλει στην πρόκληση πυρκαγιών. Για τα υπόλοιπα έτη λειτουργίας του πλοίου παρατηρείται μια συνεχής διακύμανση των ατυχημάτων ανά έτος η οποία δεν απομακρύνεται από τον μέσο όρο αισθητά για καμία ηλικία. Μία μικρή απόκλιση εμφανίζεται στο 24<sup>ο</sup> και 26<sup>ο</sup> έτος λειτουργίας του πλοίου γεγονός όμως που δεν παράγει κάποιο συμπέρασμα καθώς τα ατυχήματα αυτής της κατηγορίας είναι λίγα και δεν μπορεί να εξαχθεί κάποια τάση.



## Κεφάλαιο 4.2 Συγκεντρωτικές συχνότητες ατυχημάτων

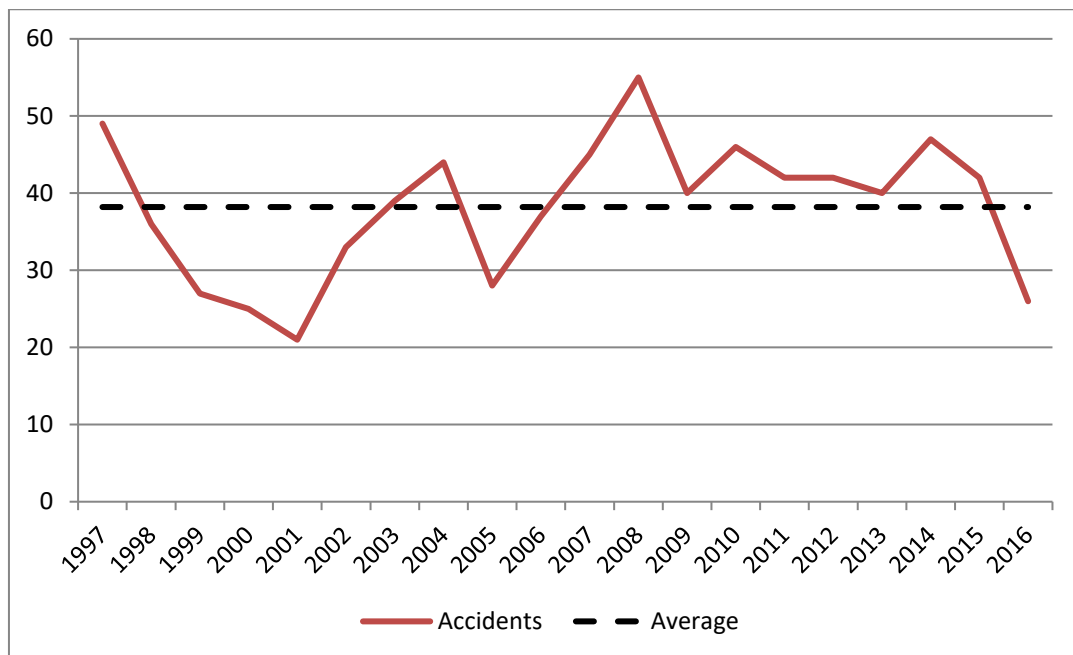
Όπως προαναφέρθηκε η μελέτη θα γίνει για ατυχήματα που συνέβησαν από το 1996 και μετά. Στον Πίνακα 11 φαίνεται η ποσοτική κατανομή των ατυχημάτων ανά τα εξεταζόμενα έτη. Στην συνέχεια στο Σχήμα 35 τα ίδια στοιχεία απεικονίζονται και σε μορφή διαγράμματος. Τα στοιχεία αυτά παρήχθησαν από την βάση δεδομένων Seaweb αθροίζοντας τον αριθμό των ατυχημάτων που συνέβησαν σε κάθε έτος.

*Πίνακας 11 Κατανομή ατυχημάτων μετά το 1996*

Year	Accidents	Percentage
1997	49	6,4%
1998	36	4,7%
1999	27	3,5%
2000	25	3,3%
2001	21	2,7%
2002	33	4,3%
2003	39	5,1%
2004	44	5,8%
2005	28	3,7%
2006	37	4,8%
2007	45	5,9%
2008	55	7,2%
2009	40	5,2%
2010	46	6,0%
2011	42	5,5%
2012	42	5,5%
2013	40	5,2%
2014	47	6,2%
2015	42	5,5%
2016	26	3,4%
Total:	764	100,0%

Από τον Πίνακα 11 και το Σχήμα 35 παρατηρείται ότι συμβαίνουν περίπου 20-60 ατυχήματα ανά έτος. Ο μέσος όρος των ατυχημάτων ανά έτος είναι 38.20 ατυχήματα. Η μέγιστη τιμή ατυχημάτων ανά έτος εμφανίστηκε το 2008 (55 ατυχήματα). Αντίστοιχα η μικρότερη τιμή παρατηρήθηκε το 2001 (21 ατυχήματα). Τα τελευταία 3 χρόνια ο αριθμός των ατυχημάτων βαίνει φθίνων πλησιάζοντας το 2016 μία από τις χαμηλότερες τιμές της τελευταίας εικοσαετίας (26 ατυχήματα)

παρά το γεγονός ότι από το 2007 έως και το 2015 ο αριθμός των ατυχημάτων ανά έτος βρισκόταν συστηματικά πάνω από τον μέσο όρο.



Σχήμα 35 Κατανομή ατυχημάτων μετά το 1996

Για τον υπολογισμό συχνοτήτων απαραίτητη προϋπόθεση είναι η γνώση των ατυχημάτων ανά έτος και ο υπολογισμός του εν λειτουργία στόλου ανά έτος (Fleet at Risk). Για τον υπολογισμό του Fleet at Risk (FaR) αρχικά έγινε μία προσπάθεια εξαγωγής του από την βάση δεδομένων Seaweb. Αν και πλήρης η εν λόγω βάση δεδομένων προσφέρει πληροφορίες μόνο για τα πλοία που είναι ενεργά την παρούσα στιγμή. Για τον υπολογισμό του Fleet at Risk για κάθε έτος αρχικά αθροίστηκε ο αριθμός των πλοίων που ήταν κατασκευασμένα μέχρι και το 1996. Στην συνέχεια στο υπολογισθέν αριθμό προστέθηκαν σταδιακά τα πλοία που κατασκευάστηκαν κάθε χρονολογία διαμορφώνοντας έτσι το Fleet at Risk κάθε έτους έως και το 2016. Γίνεται λοιπόν αντιληπτό ότι η χρήση της βάσης δεδομένων Seaweb για την εξαγωγή του Fleet at Risk θα είναι αρκετά ανακριβής για παλαιότερα έτη καθώς πλοία τα οποία ήταν ενεργά πριν από μία – δύο δεκαετίες σήμερα θα έχουν πιθανόν παροπλιστεί. Συνεπώς με την μέθοδο που χρησιμοποιήθηκε για τον υπολογισμό του ενεργού στόλου κάθε έτος αυτά τα πλοία δεν έχουν ληφθεί υπόψιν. Για την αντιμετώπιση του προβλήματος των ελλειπών στοιχείων της βάσης δεδομένων και την εξαγωγή σωστών συχνοτήτων ακολουθήθηκε η εξής διαδικασία. Τα ατυχήματα που προέκυψαν από την βάση δεδομένων Seaweb κρατήθηκαν σταθερά ενώ το Fleet at Risk της αντικαταστάθηκε

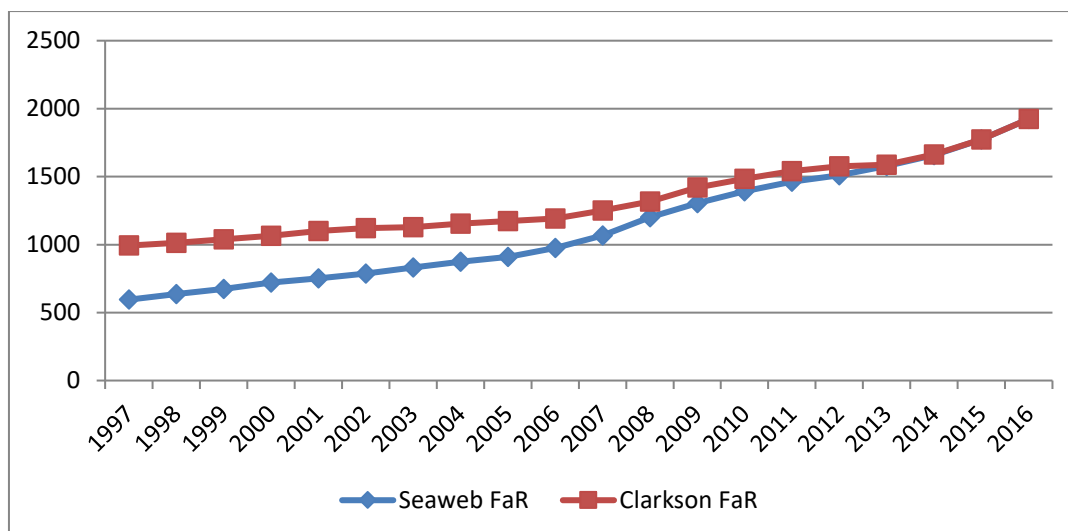
Στατιστική Ανάλυση Ναυτικών Ατυχημάτων σε πλοία LNG-LPG  
και Ανάπτυξη Αλγορίθμου Εξόρυξης Δεδομένων

από το Fleet at Risk που μας δόθηκε από την εταιρεία Clarkson (<http://www.clarksons.com/>). Στον Πίνακα 12 και στο Σχήμα 36 εμφανίζονται οι τιμές του Fleet at Risk από την Seaweb και από την Clarkson. Παρατηρείται ότι για παλαιότερα έτη η διαφορά μεταξύ των στοιχείων της Clarkson και της Seaweb είναι πολύ μεγάλη (1996: Seaweb FaR = 560 / Clarkson FaR = 952) ενώ για κοντινότερες χρονολογίες τα δύο αυτά νούμερα συγκλίνουν και τελικώς ταυτίζονται για το έτος 2016 (1924).

*Πίνακας 12 Seaweb Fleet at Risk & Frequencies*

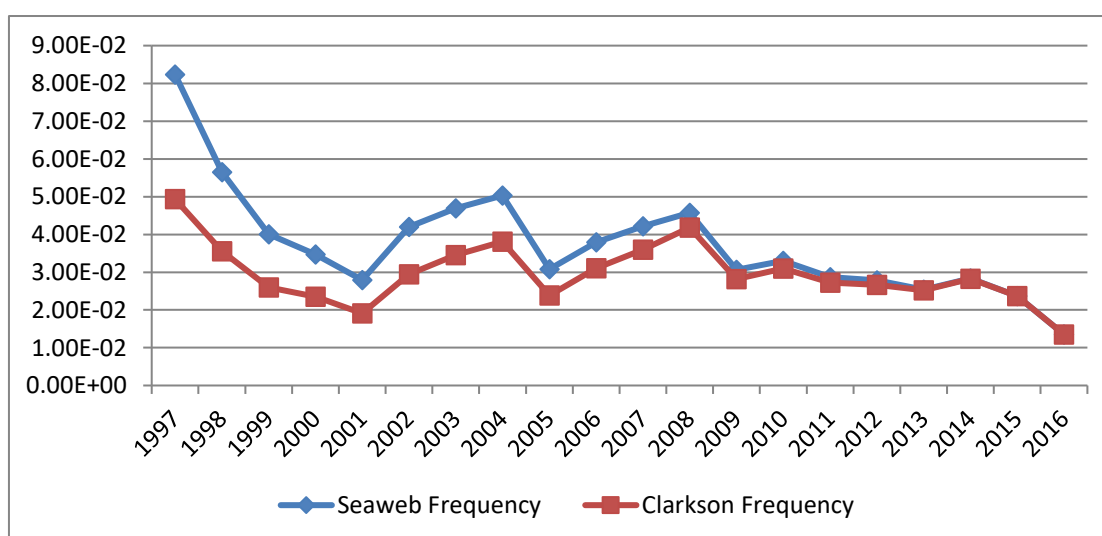
Year	Seaweb FaR	Seaweb Frequency	Clarkson FaR	Clarkson Frequency
1997	595	8,24E-02	993	4,93E-02
1998	637	5,65E-02	1012	3,56E-02
1999	674	4,01E-02	1040	2,60E-02
2000	720	3,47E-02	1065	2,35E-02
2001	752	2,79E-02	1100	1,91E-02
2002	786	4,20E-02	1121	2,94E-02
2003	832	4,69E-02	1128	3,46E-02
2004	874	5,03E-02	1155	3,81E-02
2005	909	3,08E-02	1173	2,39E-02
2006	975	3,79E-02	1191	3,11E-02
2007	1068	4,21E-02	1251	3,60E-02
2008	1202	4,58E-02	1316	4,18E-02
2009	1305	3,07E-02	1421	2,81E-02
2010	1392	3,30E-02	1484	3,10E-02
2011	1463	2,87E-02	1541	2,73E-02
2012	1509	2,78E-02	1576	2,66E-02
2013	1577	2,54E-02	1588	2,52E-02
2014	1659	2,83E-02	1663	2,83E-02
2015	1773	2,37E-02	1773	2,37E-02
2016	1924	1,35E-02	1924	1,35E-02

Στατιστική Ανάλυση Ναυτικών Ατυχημάτων σε πλοία LNG-LPG  
και Ανάπτυξη Αλγορίθμου Εξόρυξης Δεδομένων



Σχήμα 36 Σύγκριση Fleet at Risk Seaweb-Clarkson

Παρατηρείται στο Σχήμα 36 ότι ο αριθμός των πλοίων για τα έτη 1996-2016 βαίνει γνησίως αύξων το οποίο είναι αναμενόμενο για 2 λόγους. Πρώτον η ζήτηση για LNG-LPG έχει αυξηθεί την τελευταία εικοσαετία προωθώντας την κατασκευή ολοένα και περισσότερων πλοίων τέτοιου είδους. Δεύτερον όπως προαναφέρθηκε λόγω της βάσης δεδομένων Seaweb και του τρόπου λειτουργίας της πολλά πλοία τα οποία είναι παροπλισμένα σήμερα αλλά λειτουργούσαν σε παλαιότερα έτη δεν αναγράφονται στην λίστα και συνεπώς δεν συμμετέχουν στο Fleet at Risk της Seaweb που παρουσιάστηκε. Αυτός είναι και ο κύριος λόγος που υπάρχει τόσο μεγάλη διαφοροποίηση των τιμών του στόλου σε παλαιότερα έτη μεταξύ των δύο πηγών.



Σχήμα 37 Συχνότητες ατυχημάτων με χρήση Seaweb και Clarkson

Όπως φαίνεται στο Σχήμα 37 οι συχνότητες των ατυχημάτων ακολουθούν μία φθίνουσα πορεία καθώς τα ατυχήματα κυμαίνονται σταθερά μεταξύ 20-60 ανά έτος ενώ ταυτόχρονα ο αριθμός των ενεργών πλοίων ολοένα και αυξάνεται. Αν και ποσοτικά παρατηρούνται μεγάλες διακυμάνσεις μεταξύ των δύο ομάδων δεδομένων ποιοτικά δεν υπάρχουν σημαντικές διαφοροποιήσεις στην μορφή των διαγραμμάτων συχνότητων που προκύπτουν. Οι συχνότητες που προκύπτουν χρησιμοποιώντας το Fleet at Risk της Clarkson είναι πιο ακριβείς και απεικονίζουν ακριβέστερα την πραγματικότητα. Το 1997 εμφανίζεται η μέγιστη τιμή συχνότητας ατυχημάτων που είναι  $4,93E-02$  ενώ το 2016 εμφανίζεται η ελάχιστη τιμή  $1,35E-02$ . Κατά τα έτη 1997-2001 οι συχνότητες των ατυχημάτων μειώνονται σταδιακά φτάνοντας στο τοπικό ελάχιστο  $1,91E-02$  ενώ στα έτη 2002-2008 που ακολουθούν οι συχνότητες αυξάνονται σε τιμές που πλησιάζουν η και ξεπερνούν το  $4,00E-02$ . Από το 2008 έως και το 2016 οι τιμές των συχνότητων ολοένα και μειώνονται απόδειξη ότι τα ατυχήματα σε υγραεριοφόρα πλοία γίνονται ολοένα και πιο σπάνια.

## Κεφάλαιο 4.3 Συχνότητες ατυχημάτων συναρτήσει των ηλικιών

Για την εξαγωγή των συχνοτήτων των ατυχημάτων βάσει των ηλικιών των πλοίων ακολουθήθηκε η παρακάτω διαδικασία. Αρχικά υπολογίστηκε το Fleet at Risk ανά πενταετία. Οι παραχθείσες τιμές φαίνονται στον παρακάτω πίνακα και υπολογίστηκαν βάσει των στοιχείων της Clarkson τα οποία είναι και τα πιο ακριβή.

*Πίνακας 13 Αθροιστικό Fleet at Risk ανά πενταετία*

Year	FaR
1997 - 2001	5222
2002 - 2006	5781
2007 - 2011	7040
2012 - 2016	8556

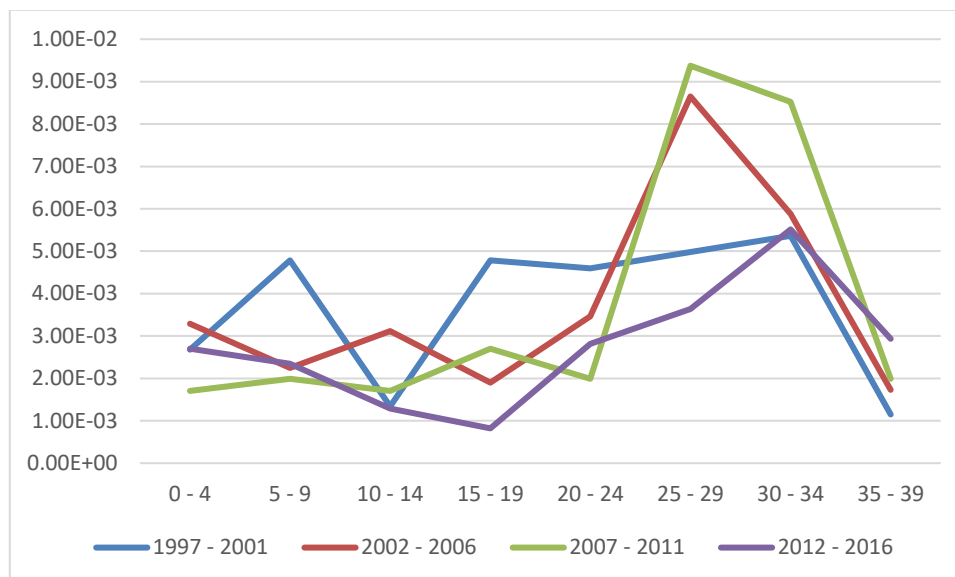
Έχοντας ορίσει το Fleet at Risk το επόμενο βήμα ήταν να υπολογιστεί ο αριθμός των ατυχημάτων που θα χρησιμοποιηθεί για τον υπολογισμό της κάθε συχνότητας. Αποφασίστηκε τα ατυχήματα τα οποία θα χρησιμοποιηθούν ως ο αριθμητής για την εξαγωγή της συχνότητας να κατηγοριοποιηθούν βάσει της ηλικίας του πλοίου. Συνεπώς τα ατυχήματα κατηγοριοποιήθηκαν βάσει 2 παραμέτρων. 1. Την χρονολογία κατά την οποία συνέβησαν 2. Την ηλικία του πλοίου όταν συνέβη το ατύχημα. Ακολουθεί ο πίνακας κατανομής τους.

*Πίνακας 14 Κατανομή ατυχημάτων ανά πενταετία συναρτήσει της ηλικίας*

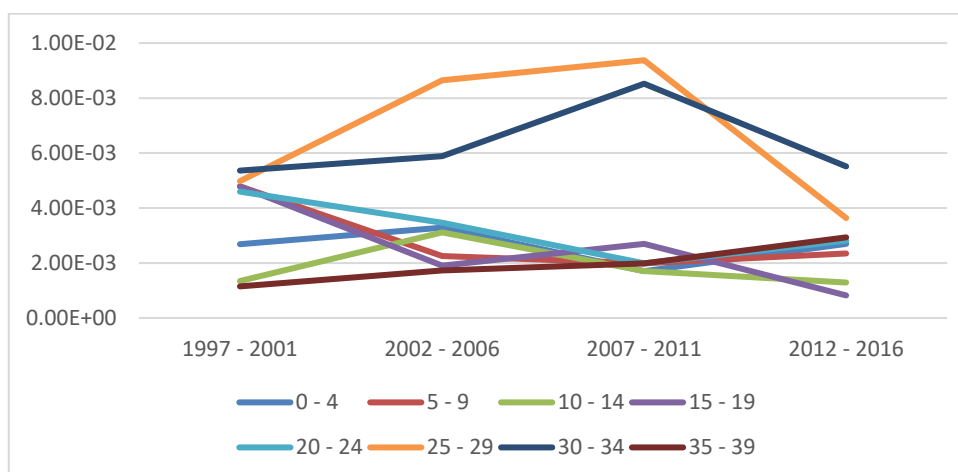
Age of Ship		Year of Accident			
From	To	1997-2001	2002-2006	2007-2011	2012-2016
0	4	14	19	12	23
5	9	25	13	14	20
10	14	7	18	12	11
15	19	25	11	19	7
20	24	24	20	14	24
25	29	26	50	66	31
30	34	28	34	60	47
35	39	6	10	14	25

Στατιστική Ανάλυση Ναυτικών Ατυχημάτων σε πλοία LNG-LPG  
και Ανάπτυξη Αλγορίθμου Εξόρυξης Δεδομένων

Χρησιμοποιώντας τα ανωτέρω υπολογισμένα Fleet at Risk και ατυχήματα ήταν δυνατόν να παραχθούν τα ακόλουθα διαγράμματα συχνοτήτων.



Σχήμα 38 Συχνότητες ατυχημάτων συναρτήσει της ηλικίας του πλοίου ανά πενταετία



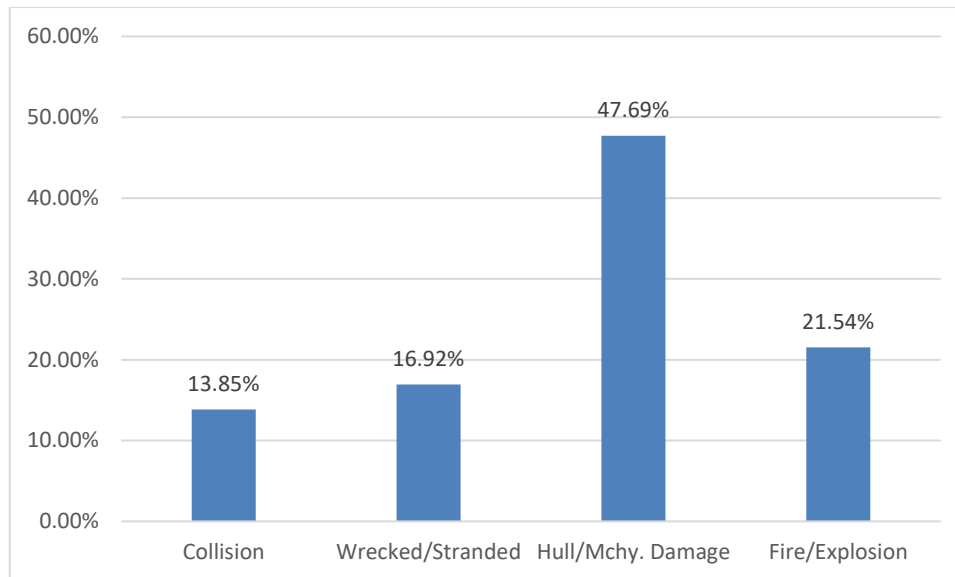
Σχήμα 39 Συχνότητες ατυχημάτων συναρτήσει της χρονολογίας ανά ηλικιακή κατηγορία

Στο Σχήμα 38 για τα έτη 1997-2001 παρατηρούμε ότι οι μέγιστες τιμές συχνοτήτων παρατηρούνται για ηλικίες 5-9 έτη και 15-34. Το Fleet at Risk για αυτά τα έτη λαμβάνεται σταθερό και ίσο με 5222. Επίσης παρατηρείται ότι για τα έτη 2002-2006 οι μέγιστες τιμές συχνοτήτων εμφανίζονται για ηλικίες 25-29 και 30-34 έτη. Το Fleet at Risk για αυτά τα έτη λαμβάνεται σταθερό και ίσο με 5781. Οι

συχρότητες των ατυχημάτων για διαφορετικές από τις προαναφερθείσες ηλικίες παραμένουν σταθερά χαμηλότερες των  $4E10^{-3}$ . Για τα έτη 2007-2011 παρατηρούμε ότι οι μέγιστες τιμές συχνοτήτων παρατηρούνται για ηλικίες 25-29 και 30-34 έτη. Το Fleet at Risk για αυτά τα έτη λαμβάνεται σταθερό και ίσο με 7040. Οι συχρότητες των ατυχημάτων για διαφορετικές από τις προαναφερθείσες ηλικίες παραμένουν σταθερά χαμηλότερες των  $3*10^{-3}$ . Στο διάγραμμα συχνοτήτων για τα έτη 2012-2016 παρατηρούμε ότι οι μέγιστες τιμές συχνοτήτων παρατηρούνται για ηλικίες 30-34 έτη. Το Fleet at Risk για αυτά τα έτη λαμβάνεται σταθερό και ίσο με 8556. Οι συχρότητες των ατυχημάτων για διαφορετικές από τις προαναφερθείσες ηλικίες παραμένουν σταθερά χαμηλότερες των  $3*10^{-3}$ . Συνολικά στο Σχήμα 38 φαίνεται ότι με την πάροδο των ετών οι μέγιστες τιμές συχνοτήτων μετατοπίζονται προς μεγαλύτερες ηλικίες πλοίων. Συγκεκριμένα ενώ για την πενταετία 1997-2001 οι συχρότητες ατυχημάτων εμφανίζουν υψηλές τιμές από 15-34 έτη στις ακόλουθες δύο πενταετίες (2002-2006 και 2007-2011) ο υψηλός αριθμός συχνοτήτων περιορίζεται σε ηλικίες 25-34 ηλικιακά έτη. Τέλος στην τελευταία πενταετία που μελετάμε 2012-2016 υψηλές τιμές συχνοτήτων εμφανίζονται στα ηλικιακά έτη 30-34. Ποσοτικά παρατηρείται ότι με την πάροδο των ετών οι υψηλές συχρότητες της τάξης του  $5E10^{-3}$  που εμφανίζονται στην πενταετία 1997-2001 σε μεγάλο ηλικιακό κομμάτι τείνουν να αντικατασταθούν στα ακόλουθα έτη από χαμηλές συχρότητες ατυχημάτων στο μεγαλύτερο κομμάτι της ζωής του πλοίου και να εμφανίζουν τοπικά ακρότατα σε συγκεκριμένες ηλικίες. Συγκεκριμένα για τις πενταετίες 2002-2006 και 2007-2011 το τοπικό ακρότατο των συχνοτήτων εμφανίζεται 25-29 ηλικιακά έτη ενώ για την τελευταία πενταετία 2012-2016 μετατοπίζεται στα 30-34 έτη. Άρα συγκεντρωτικά τα πλοία τα οποία ήταν συχνότερο να συμβεί ατύχημα στην δεκαετία 2002-2011 ήταν τα πλοία 25-29 ετών ενώ τα πλοία που είναι σήμερα συχνότερο να συμβεί ατύχημα είναι τα πλοία 30-34 ετών. Στην τελευταία πενταετία συνολικά εμφανίζονται οι χαμηλότερες τιμές συχνοτήτων. Παράλληλα στο Σχήμα 39 παρατηρούμε ότι ανεξάρτητα της χρονολογίας οι μεγαλύτερες συχρότητες ατυχημάτων εμφανίζονται σε πλοία 25-34 ετών. Στην προσπάθεια να προσεγγίσουμε τις αιτίες αυτών των υψηλών συχνοτήτων σε αυτήν την ηλικιακή κατηγορία στο Σχήμα 40 φαίνεται η κατανομή των ατυχημάτων ανά τύπο ατυχήματος για ηλικίες 25 έως 34 έτη.



Στατιστική Ανάλυση Ναυτικών Ατυχημάτων σε πλοία LNG-LPG  
και Ανάπτυξη Αλγορίθμου Εξόρυξης Δεδομένων



Σχήμα 40 Κατανομή ατυχημάτων ανά τύπο ατυχήματος για ηλικίες πλοίων 25-34 έτη

Στο Σχήμα 40 παρατηρούμε ότι περίπου τα μισά ατυχήματα που συμβαίνουν σε ηλικίες πλοίων από 25 έως 34 έτη ανήκουν στην κατηγορία Hull & Machinery Damage. Συγκρινόμενες οι ανωτέρω τιμές με τα συνολικά ατυχήματα παρατηρούμε ότι συνολικά μόλις το 35% των ατυχημάτων υπάγεται στην κατηγορία Hull & Machinery Damage. Το γεγονός αυτό είναι αναμενόμενο καθώς μετά από φθορές 25+ χρόνων όλο και περισσότερα μηχανολογικά μέρη του πλοίου αναμένεται να παρουσιάσουν κάποιο πρόβλημα ή να δυσλειτουργήσουν οδηγώντας σε κάποιο ατύχημα. Άρα για να μειωθούν οι αυξημένες συχνότητες ατυχημάτων που εμφανίζονται στις προαναφερθείσες ηλικιακές κατηγορίες θα πρέπει να γίνεται καλύτερη συντήρηση και επιθεώρηση των μηχανημάτων κατά την διάρκεια ζωής του πλοίου ώστε να αποφεύγονται αυξημένες φθορές και δυσλειτουργίες στα τελευταία έτη λειτουργίας του. Επίσης αύξηση παρατηρείται και στο ποσοστό των Fire & Explosion ατυχημάτων το οποίο από 13% συνολικά αυξάνεται στο 21.54% για πλοία ηλικίας 25 έως 34 ετών. Για τους λόγους που αναφέρθηκαν και στον προηγούμενο τύπο ατυχήματος είναι πιο πιθανή η έναρξη μίας φωτιάς ή μίας έκρηξης. Συνολικά παρατηρείται ότι η αύξηση αυτών των δύο τύπων ατυχήματος είναι αυτή που οφείλεται για τις αυξημένες συχνότητες στις ηλικίες από 25 έως 34 ηλικιακά έτη.

Η σελίδα αυτή είναι σκόπιμα κενή.

## ΚΕΦΑΛΑΙΟ 5. Εξόρυξη Δεδομένων

Σε αυτό το σημείο της διπλωματικής έχουμε εξάγει μία πληθώρα πληροφοριών για τα ατυχήματα τα είδη ατυχημάτων και τους παράγοντες που τα επηρεάζουν. Έχουμε προσδιορίσει πως επηρεάζεται κάθε είδος ατυχήματος από τους διάφορους παράγοντες που συμβάλουν σε αυτό και έχουμε στην διάθεση μας. Επίσης έχουμε εξάγει συχνότητες για τον ακριβή προσδιορισμό της μεταβολής της συχνότητας ύπαρξης ατυχημάτων ανά τα έτη. Πέραν αυτών έχουμε τελικώς προσδιορίσει και συχνότητες ύπαρξης ατυχημάτων βάσει της ηλικίας των πλοίων.

Όλα αυτά τα στοιχεία αν και χρήσιμα εκπροσωπούν το καθένα μερικώς τις αιτίες και τις τάσεις ύπαρξης ατυχημάτων. Για παράδειγμα γνωρίζουμε ότι στην περιοχή “Br. Isles N. Sea E.Chnl Biscay” έχουν συμβεί λιγότερα ατυχήματα Collision από τα αναμενόμενα ενώ τα LPG Tankers εμφανίζουν περισσότερα ατυχήματα από τα αναμενόμενα. Όμως δεν έχουμε την δυνατότητα να προσδιορίσουμε αν ένα LPG Tanker το οποίο εκτελεί μία μεταφορά στην περιοχή “Br. Isles N. Sea E.Chnl Biscay” είναι πιθανό να έχει ένα Collision ατύχημα καθώς τα επιμέρους στοιχεία είναι αντικρουόμενα και δεν γνωρίζουμε την βαρύτητα του καθενός. Οπότε η γνώση που έχουμε συγκεντρώσει μέχρι αυτό το σημείο μας δίνει την δυνατότητα να προσδιορίσουμε προβληματικές κατηγορίες πλοίων μεμονωμένα αλλά δεν μας δίνει την δυνατότητα να προσδιορίσουμε προβληματικά σενάρια μετακίνησης φορτίων με συγκεκριμένα πλοία. Σε αυτό το σημείο θα γίνει μία προσπάθεια να ξεπεραστεί αυτό το πρόβλημα με την ανάπτυξη ενός κώδικα στο λογισμικό Matlab με χρήση τεχνικών εξόρυξης δεδομένων. Συγκεκριμένα θα γίνει μία προσπάθεια ανάπτυξης ενός αλγορίθμου πρόβλεψης του τύπου ατυχήματος βάσει διαφόρων χαρακτηριστικών ενός πλοίου.

Τα τελευταία χρόνια παρατηρείται μια ταχύτατη αύξηση του όγκου των αποθηκευμένων δεδομένων. Η τάση αυτή οφείλεται στην αύξηση των δυνατοτήτων του λογισμικού και του υλικού, στη μείωση του κόστους, στη διεύρυνση της πληροφορικής σε κάθε δραστηριότητα της σύγχρονης κοινωνίας, και σε μεγάλο βαθμό στην ευρύτατη εξάπλωση του Διαδικτύου, και την σύνδεση πληθώρας συσκευών σε αυτό (σενσορες, καταγραφικά κ.α.) το οποίο καθιστά τους χρήστες πρωταγωνιστικούς παράγοντες παραγωγής πληροφοριών. Η ανάλυση όλων αυτών των δεδομένων και η ανάκτηση χρήσιμης πληροφορίας χωρίς τη χρήση εξειδικευμένων τεχνικών είναι αδύνατη. Η Εξόρυξη Δεδομένων (ΕΔ ή Data Mining), αντλώντας μεθοδολογίες από τη Μηχανική Μάθηση, τις Βάσεις Δεδομένων, τη Στατιστική και άλλους κλάδους, έχει στόχο την ανακάλυψη γνώσης μέσα από μεγάλους όγκους δεδομένων. Αν και σε επίπεδο ορολογίας η Εξόρυξη Δεδομένων χρησιμοποιείται αρκετές φορές ως συνώνυμο της Ανακάλυψης Γνώσης από Βάσεις Δεδομένων, επί της ουσίας αναγνωρίζεται ότι η Ανακάλυψη Γνώσης από Βάσεις

Δεδομένων είναι μια ευρύτερη διαδικασία και ότι η καθαυτό Εξόρυξη Δεδομένων είναι ένα στάδιο της, κατά το οποίο ανακαλύπτονται πρότυπα δεδομένων.

Το σύνολο των σταδίων της Ανακάλυψης Γνώσης από Βάσεις Δεδομένων είναι:

- α) η συλλογή, ολοκλήρωση και ο καθαρισμός των δεδομένων,
- β) η επιλογή των δεδομένων και ο μετασχηματισμός τους,
- γ) η εξόρυξη των δεδομένων και
- δ) η αξιολόγηση των σχέσεων / μοτίβων και η ανακάλυψη της γνώσης.

Υπάρχουν διάφορες εργασίες Εξόρυξης Δεδομένων και χωρίζονται σε εργασίες επιβλεπόμενης μάθησης και εργασίες μη επιβλεπόμενης μάθησης.

Οι κυριότερες εργασίες Εξόρυξης Δεδομένων είναι οι ακόλουθες:

1. Η Κατηγοριοποίηση (classification) στοχεύει στην εκτίμηση των τιμών ενός γνωρίσματος-στόχου με ονομαστικές τιμές, το οποίο ορίζει την κατηγορία των αντικειμένων.
2. Η Παλινδρόμηση (regression) μοιάζει με την κατηγοριοποίηση, αλλά το γνώρισμα-στόχος έχει αριθμητικές τιμές.
3. Η Ανάλυση Συστάδων (Clustering) επιμερίζει ένα σύνολο αντικειμένων σε ομάδες, βάσει ομοιότητας και χωρίς την ύπαρξη προκαθορισμένων κατηγοριών.
4. Η Ανάλυση Κανόνων Συσχέτισης ανακαλύπτει σχέσεις μεταξύ τιμών των γνωρισμάτων, οι οποίες εμφανίζονται συχνά μαζί.
5. Η Ανάλυση Εξαιρέσεων εντοπίζει και αναλύει περιπτώσεις, οι οποίες αποκλίνουν από το κανονικό ή συνηθισμένο.
6. Η Ανάλυση Χρονοσειρών αναλύει μεγέθη τα οποία παρουσιάζουν χρονική εξέλιξη.

Γενικά, η γνώση αποτελεί πολύτιμο κεφάλαιο και η Εξόρυξη Δεδομένων είναι το εργαλείο για την ανάκτηση της. Η Εξόρυξη Δεδομένων έχει βρει πεδίο εφαρμογής σε πλήθος διαδικασιών. Μια από αυτές είναι και η διαχείριση του ρίσκου, και με βάση αυτή την εφαρμογή θα κάνουμε χρήση στην υπάρχουσα μελέτη μας.

## Κεφάλαιο 5.1 Στάδια της Διαδικασίας Ανακάλυψης Γνώσης

Έχει επικρατήσει η πρακτική, κυρίως στη βιομηχανία λογισμικού, να θεωρούνται οι όροι «Εξόρυξη Δεδομένων» και «Ανακάλυψη Γνώσης σε Βάσεις Δεδομένων» ως συνώνυμοι, και να χρησιμοποιούνται για να περιγράψουν τη συνολική διαδικασία ανακάλυψης γνώσης. Η πρακτική αυτή ακολουθείται χάριν ευκολίας, γιατί ο όρος «Ανακάλυψη Γνώσης σε Βάσεις Δεδομένων» είναι δύσχρηστος. Επί της ουσίας όμως, επικρατεί η κοινή άποψη ότι η Ανακάλυψη Γνώσης από Βάσεις Δεδομένων (ΑΓΒΔ) είναι μια ευρύτερη διαδικασία και ότι η καθαυτό Εξόρυξη Δεδομένων αποτελεί ένα στάδιο της.

Η Ανακάλυψη Γνώσης σε Βάσεις Δεδομένων έχει ως αφετηριακό σημείο της τα πηγαία δεδομένα και ως σκοπό της την κατάλληλη επεξεργασία τους, ώστε να ανακαλυφθεί χρήσιμη γνώση. Όμως μεταξύ του αρχικού σημείου και του τελικού στόχου μεσολαβούν ενδιάμεσα στάδια, κατά τα οποία πρέπει να λάβουν χώρα συγκεκριμένες εργασίες, που επηρεάζουν σημαντικά το τελικό αποτέλεσμα. Επίσης, πρέπει να τονιστεί ότι η διαδικασία δεν είναι μονόδρομη και ότι ο αναλυτής πιθανώς θα χρειαστεί να επανέλθει σε κάποιο προηγούμενο στάδιο, να τροποποιήσει τις πρακτικές του και να επαναλάβει τα ακόλουθα στάδια. Τα στάδια της διαδικασίας Ανακάλυψης Γνώσης σε Βάσεις Δεδομένων και οι ειδικότερες εργασίες που λαμβάνουν χώρα σε κάθε ένα από αυτά είναι τα ακόλουθα:

### **Βήμα 1<sup>ο</sup> : Συλλογή, Ολοκλήρωση και Καθαρισμός των Δεδομένων.**

Αφετηριακό σημείο είναι τα πηγαία δεδομένα. Κατά κανόνα, τα δεδομένα αυτά είναι αποθηκευμένα σε διάφορες πηγές, όπως συστήματα παρακολούθησης συναλλαγών, ανεξάρτητες βάσεις δεδομένων, ανεξάρτητα αρχεία, εξωτερικές πηγές κλπ. Επίσης, τα δεδομένα αυτά πάσχουν από διάφορα προβλήματα όπως σφάλματα, αντιφάσεις, χαμένες τιμές κλπ. Τα αρχικά δεδομένα πρέπει να συλλεχτούν από τις διάφορες πηγές, να ομογενοποιηθούν και να καθαριστούν. Στο σημείο αυτό οφείλουμε να τονίσουμε τη σημασία της ποιότητας των δεδομένων. Προβληματικά δεδομένα, τα οποία περιέχουν εσφαλμένες, ακραίες ή χαμένες τιμές μπορεί να αποπροσανατολίσουν τους αλγόριθμους εξόρυξης και να οδηγήσουν στην εξαγωγή άκυρων και εσφαλμένων σχέσεων και μοτίβων. Ορισμένοι αλγόριθμοι εξόρυξης δεδομένων έχουν τη δυνατότητα να αντιμετωπίζουν ενδογενώς τα προβλήματα των δεδομένων, ωστόσο αυτό δεν ισχύει γενικώς και επίσης, ο τρόπος αντιμετώπισης δεν είναι πάντα ο καλύτερος. Για τον λόγο αυτό, είναι προτιμότερο ο καθαρισμός των δεδομένων να γίνει από τον αναλυτή ως ανεξάρτητη εργασία και με τρόπο ελεγχόμενο από αυτόν. Συνήθως, τα δεδομένα,

αφού απαλλαγούν από τα προβλήματα τους, αποθηκεύονται σε μια Αποθήκη Δεδομένων.

### **Βήμα 2<sup>ο</sup> : Επιλογή Δεδομένων και Μετασχηματισμός τους.**

Για να διεξαχθεί επιτυχώς η Εξόρυξη Δεδομένων, πρέπει καταρχήν να δημιουργηθεί το κατάλληλο σύνολο δεδομένων (data set). Οι μέθοδοι της ΕΔ είναι ισχυρώς καθοδηγούμενες από τα δεδομένα (data driven). Αυτό σημαίνει ότι το όποιο αποτέλεσμα θα αντληθεί απευθείας από τα δεδομένα. Για τον λόγο αυτό, η επιλογή των κατάλληλων δεδομένων είναι κομβικής σημασίας.

Επιλογή δεδομένων σημαίνει καταρχάς επιλογή των κατάλληλων γνωρισμάτων ή χαρακτηριστικών (attributes– features). Με όρους πίνακα, αυτό θα σήμαινε επιλογή των κατάλληλων στηλών. Η επιλογή των γνωρισμάτων είναι άμεσα συναρτημένη με την εργασία που εκτελεί ο αναλυτής. Ορισμένα γνωρίσματα μπορεί να είναι χρήσιμα για μια εργασία, ενώ ορισμένα άλλα για κάποια άλλη. Ο αναλυτής αρχικά επιλέγει τα γνωρίσματα τα οποία θεωρεί ότι περιέχουν ουσιαστική πληροφορία που σχετίζεται με την ανάλυση του. Σε ορισμένες περιπτώσεις πρέπει να γίνει και επιλογή παραδειγμάτων, αντικειμένων ή γραμμών του πίνακα. Η εργασία αυτή ονομάζεται δειγματοληψία και γίνεται όταν τα δεδομένα είναι πάρα πολλά και οι αλγόριθμοι επεξεργασίας έχουν πρόβλημα με τον χειρισμό του όγκου τους.

Η αρχική υποκειμενική επιλογή χαρακτηριστικών δεν είναι αρκετή. Στο αρχικό στάδιο, ο αναλυτής αποκλείει τα γνωρίσματα που εμφανώς δεν σχετίζονται με την ανάλυση του. Στη συνέχεια όμως, η επιλογή μπορεί να μην είναι προφανής, καθώς το ίδιο μέγεθος μπορεί να καταγράφεται με διαφορετικούς τρόπους. Δεδομένου ότι αναλύονται πολλά στοιχεία, κάθε ένα από τα οποία έχει καταγραφεί με πολλούς τρόπους, καθώς και ότι μπορεί να υπάρχουν σχέσεις αλληλεξάρτησης μεταξύ των γνωρισμάτων, το πρόβλημα περιπλέκεται.

Τίθεται λοιπόν το ερώτημα ποιο είναι το βέλτιστο υποσύνολο γνωρισμάτων, το οποίο είναι το πλέον κατάλληλο για τη συγκεκριμένη εργασία που εκτελεί ο αναλυτής. Η εργασία επιλογής στηλών ονομάζεται επιλογή χαρακτηριστικών (feature selection).

Στο στάδιο αυτό γίνεται και ο μετασχηματισμός των δεδομένων. Για παράδειγμα, μπορεί να γίνει αναγωγή αριθμητικών τιμών σε άλλες αριθμητικές τιμές ή μετατροπή αριθμητικών τιμών σε ονομαστικές τιμές. Τέτοιες εργασίες γίνονται συνήθως για να προσαρμοστούν τα δεδομένα σε απαιτήσεις των μεθόδων ανάλυσης. Για παράδειγμα, ορισμένες μέθοδοι κατηγοριοποίησης επηρεάζονται ισχυρά από πεδία με μεγάλες τιμές και ασθενώς από πεδία με μικρές τιμές. Σε τέτοιες περιπτώσεις, οι τάξεις μεγέθους των τιμών στα διάφορα πεδία πρέπει να

είναι συγκρίσιμες. Το τελικό αποτέλεσμα αυτού του σταδίου είναι ένα σύνολο δεδομένων που θα χρησιμοποιηθεί για την εξαγωγή μοτίβων.

### **Βήμα 3<sup>ο</sup> : Εξόρυξη Δεδομένων.**

Στο στάδιο αυτό γίνεται η καθαυτό Εξόρυξη Δεδομένων, δηλαδή η εξαγωγή μοτίβων. Αρχικά ο αναλυτής πρέπει να επιλέξει το είδος εργασίας Εξόρυξης Δεδομένων που θα εφαρμόσει. Σε γενικές γραμμές, ο χρήστης μπορεί να εκτελέσει Περιγραφική Ανάλυση (descriptive analytics) ή Προγνωστική Ανάλυση (predictive analytics). Η Περιγραφική Ανάλυση στοχεύει στην κατάδειξη ομαδοποιήσεων και ιδιοτήτων των δεδομένων, χωρίς να επιδιώκει τη διατύπωση προβλέψεων. Αντιθέτως, η προγνωστική ανάλυση στοχεύει στη διατύπωση προβλέψεων για το μέλλον, συνήθως με την οικοδόμηση κάποιου μοντέλου. Υπάρχουν διάφορα είδη εργασιών ΕΔ, όπως η κατηγοριοποίηση, η ανάλυση συστάδων και η ανάλυση κανόνων συσχέτισης. Αναλόγως με το πρόβλημα που θέλει να μελετήσει, ο αναλυτής επιλέγει το είδος εργασίας ΕΔ.

Επιπλέον, για κάθε είδος εργασίας υπάρχουν πολλές μέθοδοι. Για παράδειγμα, κατηγοριοποίηση μπορεί να γίνει με χρήση Νευρωνικών Δικτύων, Δένδρων Αποφάσεων, Μπαϋεσιανών Κατηγοριοποιητών κλπ. Κάθε μια από αυτές τις μεθόδους πλεονεκτεί σε ορισμένα θέματα έναντι των άλλων και μειονεκτεί σε άλλα. Ο αναλυτής, ανάλογα με τις προτεραιότητες του, θα επιλέξει τη μέθοδο που θα χρησιμοποιήσει. Επίσης, είναι πιθανό να χρησιμοποιήσει περισσότερες από μια μεθόδους και να συγκρίνει τα αποτελέσματα. Κατά τη διάρκεια της εφαρμογής των μεθόδων, ο αναλυτής θα χρειαστεί να ρυθμίσει τις παραμέτρους τους.

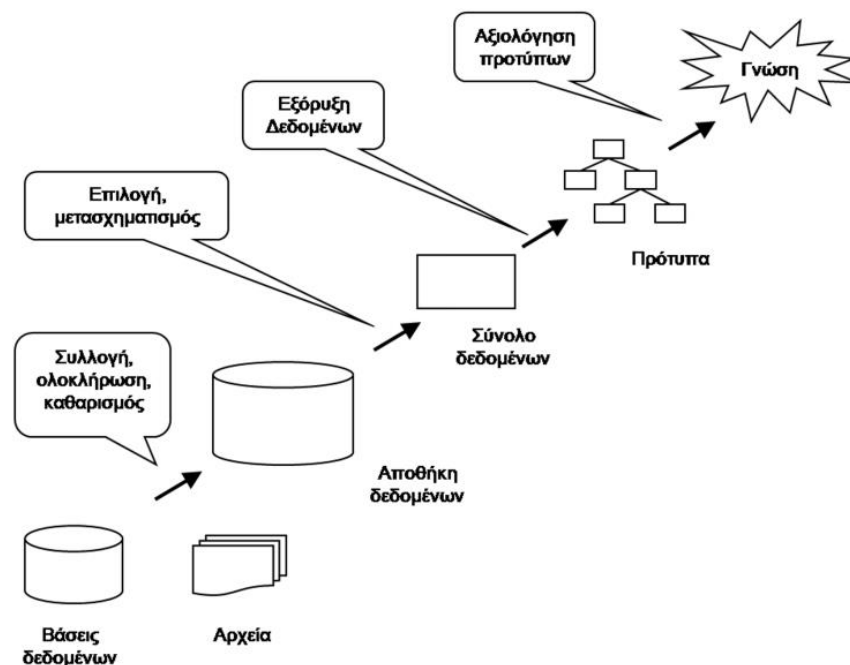
### **Βήμα 4<sup>ο</sup> : Αξιολόγηση Μοτίβων και Ανακάλυψη Γνώσης.**

Ο αναλυτής αξιολογεί τα μοτίβα δεδομένων που προέκυψαν από το προηγούμενο στάδιο. Εάν τα αποτελέσματα δεν είναι ικανοποιητικά θα επανέλθει σε προηγούμενα στάδια και θα επαναλάβει τις εργασίες. Πιθανώς θα χρειαστεί να τροποποιήσει το σύνολο δεδομένων ή να χρησιμοποιήσει διαφορετικές μεθόδους ΕΔ. Για την αναπαράσταση της γνώσης μπορούν να εφαρμοστούν τεχνικές οπτικοποίησης. Με την ολοκλήρωση αυτού του σταδίου ο αναλυτής έχει εξάγει τα συμπεράσματα του και έπεται η λήψη αποφάσεων και η ανάληψη δράσης.

### **Βήμα 5<sup>ο</sup> : Εργασίες Εξόρυξης Δεδομένων**

Στο στάδιο της Εξόρυξης Δεδομένων ο αναλυτής εξάγει πρότυπα από το σύνολο δεδομένων που έχει δημιουργήσει. Ανάλογα με το είδος του προβλήματος και τον τύπο της ανάλυσης που θα διεξάγει, μπορεί να αναζητήσει πρότυπα διάφορων τύπων. Υπάρχουν ποικίλες εργασίες Εξόρυξης Δεδομένων. Οι εργασίες αυτές μπορούν να ταξινομηθούν με διάφορους τρόπους. Ένας διαχωρισμός των μεθόδων ΕΔ είναι σε μεθόδους επιβλεπόμενης μάθησης (supervised learning) και μεθόδους

μη επιβλεπόμενης μάθησης (unsupervised learning). Η επιβλεπόμενη μάθηση έχει στόχο τη μοντελοποίηση των σχέσεων ανάμεσα σε ένα εξαρτημένο γνώρισμα – στόχο και σε άλλα ανεξάρτητα γνωρίσματα. Η ανάλυση συνίσταται στην τυποποίηση των σχέσεων ανάμεσα στην εξαρτημένη και στις ανεξάρτητες μεταβλητές, συνήθως με τη δημιουργία ενός μοντέλου, που επιτρέπει τον υπολογισμό της εξαρτημένης μεταβλητής από τις ανεξάρτητες. Το μοντέλο μπορεί να χρησιμοποιηθεί για τη διατύπωση προβλέψεων. Το όνομα «επιβλεπόμενη μάθηση» σημαίνει ότι το γνώρισμα στόχος και οι τιμές του καθοδηγούν τη διαδικασία μάθησης. Στη μη επιβλεπόμενη μάθηση δεν υπάρχει κάποια στήλη στόχος και οι αλγόριθμοι προσπαθούν να ομαδοποιήσουν τα δεδομένα σε ομάδες που δεν είναι γνωστές εκ των προτέρων. Στην Εικόνα 7 απεικονίζονται τα στάδια της διαδικασίας ανακάλυψης γνώσης.



Εικόνα 7 Στάδια της διαδικασίας Ανακάλυψης γνώσης (Πηγή <http://slideplayer.gr>)



## Κεφάλαιο 5.2 Εξόρυξη Δεδομένων Εργασίες και Αλγόριθμοι

### Κεφάλαιο 5.2.1 Ορισμός

Η Εξόρυξη Δεδομένων αποτελεί την απάντηση στο ερώτημα – Έχουμε γενικά μια πληθώρα δεδομένων, όμως που και ποια είναι η πληροφορία που μπορούμε να εξάγουμε από αυτά. Ο ανθρώπινος νους έχει περιορισμένες αναλυτικές δυνατότητες. Ακόμα και χωρίς την εκρηκτική αύξηση του όγκου των δεδομένων που παρατηρήθηκε τα τελευταία χρόνια, τού είναι πολύ δύσκολο να επεξεργαστεί αποτελεσματικά τα διαθέσιμα δεδομένα. Η επεξεργασία των στοιχείων αυτών χωρίς εξειδικευμένα εργαλεία, αν δεν είναι αδύνατη, είναι αργή, ακριβή και εν πολλοίς υποκειμενική. Η επιστήμη της Στατιστικής προσφέρει λύσεις ανάλυσης δεδομένων, δεν αντιμετωπίζει όμως ικανοποιητικά την επεξεργασία μεγάλου όγκου πληροφοριών. Επίσης η Μηχανική Μάθηση και η Αναγνώριση Μοτίβων διαθέτουν τις δικές τους μεθοδολογίες, όμως και πάλι δεν αντιμετωπίζουν το πρόβλημα του όγκου των δεδομένων. Ο κλάδος των Βάσεων Δεδομένων είναι ο κατ' εξοχήν αρμόδιος για την τήρηση μεγάλου όγκου δεδομένων, όμως η σχεδιαστική φιλοσοφία του είναι προσανατολισμένη στην καταχώρηση, στη διαχείριση και στην ανάκτηση των δεδομένων, όχι όμως και στην ανάλυση τους. Η Εξόρυξη Δεδομένων αποτελεί αποτέλεσμα της ανάγκης για επεξεργασία των αποθηκευμένων δεδομένων και εξαγωγή χρήσιμης πληροφορίας. Αντλώντας μεθοδολογίες από όλους τους επιστημονικούς κλάδους που αναφέρθηκαν παραπάνω, καθώς και από άλλους, όπως η Οπτικοποίηση, στοχεύει στην ανακάλυψη πολύτιμης γνώσης, που είναι κρυμμένη σε μεγάλους όγκους δεδομένων.

Η Εξόρυξη Δεδομένων (ΕΔ) (Data Mining (DM)) ορίζεται ως η διαδικασία ανακάλυψης σχέσεων και μοτίβων μέσα από δεδομένα, δίνοντας έτσι έμφαση στη διάσταση της Μηχανικής Μάθησης[1]. Η Εξόρυξη Δεδομένων συνίσταται στην ανακάλυψη ή «εξόρυξη» γνώσης από μεγάλους όγκους δεδομένων[2]. Ο ορισμός αυτός τονίζει τη διάσταση του όγκου των δεδομένων. Άλλοι συγγραφείς, χρησιμοποιούν τον όρο Ανακάλυψη Γνώσης σε Βάσεις Δεδομένων (Knowledge Discovery in Databases – KDD) για τη συνολική διαδικασία ανακάλυψης σχέσεων μέσα από μεγάλα και περίπλοκα σύνολα δεδομένων [3]. Σύμφωνα με το σκεπτικό αυτό, η ανακάλυψη γνώσης από τα δεδομένα συνίσταται σε μια διαδικασία, που ξεκινά από τα πηγαία δεδομένα και καταλήγει στην τελική διατύπωση συμπερασμάτων και στη λήψη αποφάσεων, μέσα από μια αλληλουχία διαδοχικών σταδίων. Η καθαυτό Εξόρυξη Δεδομένων αποτελεί ένα από τα στάδια αυτής της διαδικασίας και συνίσταται στον σκληρό πυρήνα της. Περιλαμβάνει την εφαρμογή αλγορίθμων και την κατασκευή μοντέλων, τα οποία στοχεύουν στην ανακάλυψη και εξαγωγή μοτίβων (patterns). Η συνολική διαδικασία ανακάλυψης γνώσης καλείται

να αντιμετωπίσει προβλήματα, όπως το πώς είναι αποθηκευμένα τα δεδομένα και πως επιτυγχάνεται η πρόσβαση σε αυτά, πως οι αλγόριθμοι εξαγωγής σχέσεων και μοτίβων πρέπει να κλιμακωθούν, ώστε να είναι ικανοί να χειριστούν τον όγκο των δεδομένων, πως θα γίνει η οπτικοποίηση των αποτελεσμάτων, ώστε να καταστούν κατανοητά κλπ.

Ένα ζήτημα που οφείλει να γίνει κατανοητό, είναι το τι σημαίνει «ανακάλυψη γνώσης», δηλαδή τι τελικά εξάγεται. Από τεχνικής σκοπιάς αυτό που τελικά εξάγεται είναι κανονικότητες και πρότυπα δεδομένων, που περιγράφουν ή διαφοροποιούν κατηγορίες ή περιπτώσεις και που πιθανώς να χρησιμεύσουν για τη διατύπωση προβλέψεων. Η αξιολόγηση αυτών των σχέσεων και η ερμηνεία τους συνιστά τη γνώση, όπως αυτή γίνεται αντιληπτή από τον άνθρωπο.

## Κεφάλαιο 5.2.2 Μηχανισμοί Εξόρυξης Δεδομένων

### Κατηγοριοποίηση(Classification).

Η κατηγοριοποίηση είναι μια από τις συνηθέστερες εργασίες Εξόρυξης Δεδομένων. Πρόκειται για εργασία επιβλεπόμενης μάθησης. Σε προβλήματα Κατηγοριοποίησης είναι γνωστό εκ των προτέρων ότι τα δεδομένα υπάγονται σε κατηγορίες. Σε ένα από τα γνωρίσματα καταγράφεται η κατηγορία των αντικειμένων.

Το μοντέλο δεν είναι υποχρεωτικά κανόνες, αλλά μπορεί να έχει άλλες μορφές, όπως πχ να συνίσταται σε ένα πλέγμα κόμβων και συνδέσεων ενός νευρωνικού δικτύου. Αφού δημιουργηθεί το μοντέλο, μπορεί να χρησιμοποιηθεί για τη διατύπωση προβλέψεων.

### Παλινδρόμηση(Regression).

Η παλινδρόμηση είναι μια εργασία επιβλεπόμενης μάθησης που μοιάζει πολύ με την κατηγοριοποίηση. Υπάρχει πάλι ένα γνώρισμα-στόχος, οι τιμές του οποίου υπολογίζονται από τα υπόλοιπα γνωρίσματα. Οι αλγόριθμοι παλινδρόμησης εξετάζουν τις σχέσεις μεταξύ του γνωρίσματος-στόχου και των υπόλοιπων γνωρισμάτων και κατασκευάζουν έναν μηχανισμό υπολογισμού. Η διαφορά με την κατηγοριοποίηση είναι ότι στην περίπτωση της παλινδρόμησης υπολογίζονται αριθμητικές τιμές.

### Ανάλυση Συστάδων(Clustering).

Η Ανάλυση Συστάδων είναι και αυτή μια πολύ συνηθισμένη εργασία Εξόρυξης Δεδομένων και ανήκει στην κατηγορία εργασιών μη επιβλεπόμενης μάθησης. Στόχος της Ανάλυσης Συστάδων είναι ο επιμερισμός ενός συνόλου αντικειμένων σε ομάδες. Η διαφορά με την Κατηγοριοποίηση έγκειται στο γεγονός ότι δεν υπάρχουν κατηγορίες, οι οποίες είναι εκ των προτέρων γνωστές, δεν υπάρχει δηλαδή ένα γνώρισμα στο οποίο καταγράφεται η κατηγορία των αντικειμένων. Οι αλγόριθμοι της Ανάλυσης Συστάδων προσπαθούν να δημιουργήσουν ομάδες αναζητώντας ομοειδή αντικείμενα. Στόχος τους είναι να μεγιστοποιήσουν την ομοιότητα εντός των ομάδων και την ανομοιότητα μεταξύ των ομάδων. Αυτό σημαίνει ότι τα αντικείμενα της μιας ομάδας πρέπει να μοιάζουν μεταξύ τους και να μην μοιάζουν με τα αντικείμενα των άλλων ομάδων. Αφού σχηματιστούν οι ομάδες, μπορούν να θεωρηθούν ως κατηγορίες και να δημιουργηθούν κανόνες που να τις περιγράφουν

### **Ανάλυση Κανόνων Συσχέτισης(Link Analysis).**

Η Ανάλυση Κανόνων Συσχέτισης θεωρείται το πιο γνήσιο παράγωγο της Εξόρυξης Δεδομένων, καθώς οι άλλες μέθοδοι προέρχονται από τη Μηχανική Μάθηση, τη Στατιστική κλπ. Στόχος των Κανόνων Συσχέτισης είναι η ανακάλυψη σχέσεων μεταξύ τιμών των γνωρισμάτων, οι οποίες εμφανίζονται συχνά μαζί. Για την καλύτερη κατανόηση του αντικειμένου παραθέτουμε το παρακάτω παράδειγμα.

### **Ανάλυση Εξαιρέσεων.**

Όλες οι εργασίες που αναφέρθηκαν μέχρι τώρα αφορούν τη διατύπωση «κανόνων γενικής ισχύος», την τυποποίηση δηλαδή σχέσεων που αφορούν μεγάλες ομάδες αντικειμένων. Στις εργασίες αυτές ιδιόμορφα και σπάνια γεγονότα απορρίπτονται ως μη χρήσιμη πληροφορία. Υπάρχουν όμως περιπτώσεις, που το ενδιαφέρον βρίσκεται ακριβώς στις εξαιρέσεις. Οι μέθοδοι Ανάλυσης Εξαιρέσεων επιχειρούν να εντοπίσουν τις εξαιρέσεις, χρησιμοποιώντας στατιστικές κατανομές πιθανοτήτων ή μέτρα απόστασης που βασίζονται στην ομοιότητα. Η μη ομαλή συμπεριφορά διαφόρων δεικτών και ποσοτήτων μπορεί να σηματοδοτεί προβλήματα σε διάφορες λειτουργίες. Ο άμεσος εντοπισμός τέτοιων ανωμαλιών επιτρέπει την έγκαιρη διάγνωση του προβλήματος και την ανάληψη κατάλληλης δράσης για την αντιμετώπιση του προβλήματος.

Παρακάτω θα αναφέρουμε το θεωρητικό υπόβαθρο είτε αναλυτικά είτε ποιοτικά (λόγο πολυπλοκότητας που παρουσιάζουν και δεν είναι αντικείμενο της παρούσας εργασίας η αναλυτική παρουσίαση τους) με αναφορές στη διεθνή βιβλιογραφία ώστε να μπορεί να ανατρέξει ο αναγνώστης, για τους αλγορίθμους εξόρυξης δεδομένων που χρησιμοποιήσαμε εδώ.

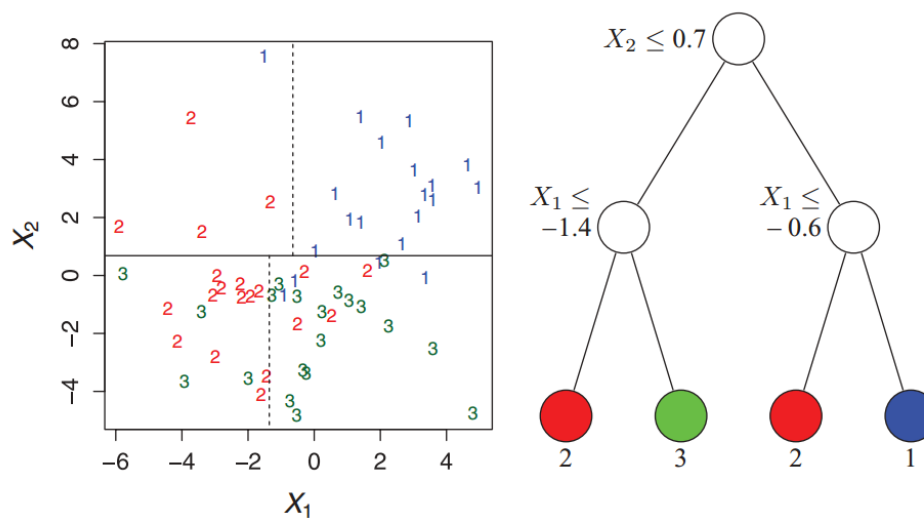
## Κεφάλαιο 5.2.3 Κατηγορίες Αλγορίθμων Εξόρυξης Δεδομένων

### ΔΕΝΤΡΑ ΚΑΤΗΓΟΡΙΟΠΟΙΗΣΗΣ

Σε ένα πρόβλημα κατηγοριοποίησης, έχουμε ένα δείγμα  $n$  παρατηρήσεων σε μια μεταβλητή κατηγορίας  $Y$  που παίρνει τις τιμές  $1, 2, \dots, k$ , και  $p$  μεταβλητές πρόγνωσης  $X_1, \dots, X_p$ . Ο στόχος μας είναι να βρούμε ένα μοντέλο για την πρόβλεψη των τιμών του  $Y$  από τις νέες τιμές του  $X$ . Θεωρητικά, η λύση είναι απλά μια διαμέριση του χώρου της  $X$  σε  $k$  ξένα μεταξύ τους σύνολα,  $A_1, A_2, \dots, A_k$ , έτσι ώστε η προβλεφθείσα τιμή του  $Y$  είναι  $j$  εάν το  $X$  ανήκει στο  $A_j$  για  $j = 1, 2, \dots, k$ . Εάν οι μεταβλητές  $X$  παίρνουν διαταγμένες τιμές, δύο κλασσικές λύσεις είναι η γραμμική ανάλυση διακρίσεως (linear discriminant analysis) [4] και η κατηγοριοποίηση του κοντινότερου γείτονα [5]. Αυτές οι μέθοδοι παράγουν σύνολα  $A_j$  με κατά τμήματα γραμμικά και μη γραμμικά, αντίστοιχα, όρια που δεν είναι εύκολο να ερμηνευθούν εάν το  $p$  είναι μεγάλο.

Οι μέθοδοι δέντρων κατηγοριοποίησης παράγουν τα ορθογώνια σύνολα  $A_j$  διαμερίζοντας αναδρομικά το σύνολο των δεδομένων ανά μια μεταβλητή  $X$  κάθε φορά. Αυτό καθιστά τα σύνολα ευκολότερα στο να ερμηνευτούν. Παραδείγματος χάριν, το σχήμα 1 δίνει ένα παράδειγμα όπου υπάρχουν τρεις κατηγορίες και δύο μεταβλητές  $X$ . Στο αριστερό τμήμα σχεδιάζονται τα σημεία των δεδομένων και οι διαμερίσεις ενώ το δεξί παρουσιάζει την αντίστοιχη δομή του δέντρου απόφασης. Ένα βασικό πλεονέκτημα της δομής των δέντρων είναι η δυνατότητα εφαρμογής της σε οποιοδήποτε αριθμό μεταβλητών, ενώ το γράφημα στα αριστερά της περιορίζεται το πολύ σε δύο.

Στατιστική Ανάλυση Ναυτικών Ατυχημάτων σε πλοία LNG-LPG  
και Ανάπτυξη Αλγορίθμου Εξόρυξης Δεδομένων



Εικόνα 8 Διαμερίσεις (αριστερά) και δομή δέντρων απόφασης (δεξιά) για ένα πρότυπο δέντρο κατηγοριοποίησης με τρεις κατηγορίες που ονομάζουμε 1, 2, και 3. Σε κάθε ένα ενδιάμεσο κόμβο, μια περίπτωση πηγαίνει στον αριστερό κόμβο απογόνων εάν και μόνο εάν η συνθήκη ικανοποιείται. Η προβλεπόμενη κατηγορία δίνεται κάτω από κάθε φυλλοειδή κόμβο (Πηγή <http://slideplayer.gr>)

Ο πρώτος δημοσιευμένος αλγόριθμος δέντρων κατηγοριοποίησης είναι ο THAID [5,6] Υιοθετώντας ένα μέτρο της μη καθαρότητας των κόμβων βασισμένο στη κατανομή των παρατηρηθεισών τιμών της  $Y$  στον κόμβο, ο THAID διαμερίζει ένα κόμβο με εξάντληση αναζήτησης πάνω σε κάθε  $X$  και κάθε  $S$  για τη διαμέριση  $\{X \in S\}$  ώστε να ελαχιστοποιεί τη συνολική μη καθαρότητα των δυο κόμβων απογόνων.

Εάν το  $X$  παίρνει διαταγμένες τιμές, το σύνολο  $S$  είναι ένα διάστημα της μορφής  $(-\infty, c]$ . Διαφορετικά, το  $S$  είναι ένα υποσύνολο των τιμών που λαμβάνονται από το  $X$ . Η διαδικασία εφαρμόζεται κατ' επανάληψη στα στοιχεία σε κάθε κόμβο απογόνων. Η διαμέριση σταματά εάν η σχετική μείωση στην μη καθαρότητα είναι κάτω από το προκαθορισμένο κατώτατο όριο..

### Δέντρα παλινδρόμησης

Ένα δέντρο παλινδρόμησης είναι παρόμοιο με ένα δέντρο κατηγοριοποίησης, εκτός από το ότι η μεταβλητή  $Y$  παίρνει διαταγμένες τιμές και ένα μοντέλο παλινδρόμησης εγκαθίσταται σε κάθε κόμβο για να δώσει τις προβλεφθείσες τιμές του  $Y$ . Ιστορικά, ο πρώτος αλγόριθμος δέντρων παλινδρόμησης είναι ο AID, [7] που εμφανίστηκε αρκετά χρόνια πριν από τον THAID. Οι μέθοδοι δέντρων παλινδρόμησης AID και CART ακολουθούν τον αλγόριθμο 1, με την ακαθαρσία των κόμβων να είναι το άθροισμα των τετραγωνικών αποκλίσεων για το μέσο όρο και τον κόμβο που προβλέπουν το μέσο δειγματικό όρο του  $Y$ . Αυτό παράγει κατά

τμήματα τα σταθερά μοντέλα. Αν και είναι απλό να ερμηνευθεί, η ακρίβεια πρόβλεψης αυτών των μοντέλων υστερεί πολλές φορές σε σχέση με αυτών των μοντέλων με περισσότερη ομαλότητα. Ωστόσο μπορεί να είναι υπολογιστικά ανέφικτο, να επεκταθεί αυτή η προσέγγιση στα κατά τμήματα γραμμικά μοντέλα, επειδή δύο γραμμικά μοντέλα (ένα για κάθε κόμβο απόγονων) πρέπει να προσαρμοστεί για κάθε υποψήφια διάσπαση. Ο M5' μια προσαρμογή ενός αλγορίθμου δέντρων παλινδρόμησης που προτάθηκε από τον Quinlan [8] χρήσις μια υπολογιστικά πιο αποδοτική στρατηγική για να κατασκευάσει τα κατά τμηματικά τα γραμμικά μοντέλα. Κατασκευάζει αρχικά ένα τμηματικά σταθερό δέντρο και μετά προσαρμόζει ένα μοντέλο γραμμικής παλινδρόμησης στα δεδομένα σε κάθε φυλλοειδή κόμβο. Επειδή η δομή δέντρων είναι η ίδια με αυτήν ενός κατά τμήματα σταθερού μοντέλου, τα προκύπτοντα δέντρα τείνουν να είναι μεγαλύτερα από εκείνα που προκύπτουν από άλλες τμηματικά γραμμικές μεθόδους κατασκευής δέντρων.

Ο GUIDE [9] χρησιμοποιεί τεχνικές δέντρων κατηγοριοποίησης για να επιλύσει το πρόβλημα της παλινδρόμησης. Σε κάθε κόμβο, εγκαθιστά ένα μοντέλο παλινδρόμησης στα δεδομένα και υπολογίζει τα υπόλοιπα(residuals). Κατόπιν καθορίζει μια κατηγορία μεταβλητής Y' με τιμές 1 ή 2, ανάλογα με εάν το πρόσημο του υπολοίπου είναι θετικό ή όχι. Τέλος, εφαρμόζει τον αλγόριθμο 2 στη μεταβλητή Y' για να χωρίσει τον κόμβο σε δύο τμήματα. Αυτή η προσέγγιση έχει τρία πλεονεκτήματα: (1) οι διασπάσεις είναι αμερόληπτες, (2) μόνο ένα μοντέλο παλινδρόμησης προσαρμόζεται σε κάθε κόμβο και (3) επειδή είναι βασισμένη στα υπόλοιπα(residuals), η μέθοδος δεν περιορίζεται ούτε στα κατά τμήματα σταθερά μοντέλα ούτε στο κριτήριο ελαχίστων τετραγώνων. Ο Πίνακας 15 συνοψίζει τα κύρια χαρακτηριστικά γνωρίσματα των αλγορίθμων κατασκευής δέντρων παλινδρόμησης GUIDE, CART, και M5'.

*Πίνακας 15 Σύγκριση αλγορίθμων δέντρων παλινδρόμησης (με tick δηλώνεται ότι υποστηρίζεται το χαρακτηριστικό).*

Χαρακτηριστικά	CART	GUIDE	M5'
Αμερόληπτη Διαμέριση		✓	
Είδος Διαμέρισης	$u, l$	$u$	$u$
Κλάδοι/Διαμέριση	2	2	$\geq 2$
Δοκιμές αλληλεπίδρασης		✓	
Απομείωση (Κλαδεμα)	✓	✓	✓
Σειρά Κατάταξης Μεταβλητών	✓	✓	
Μοντέλα Κόμβων	$c$	$c, m, p, r$	$c, r$
Μέθοδοι Χαμένων τιμών	$s$	$a$	$g$
Κριτήρια απωλειών	$v$	$v, w$	$v$
Bagging & Ensembles		✓	

### Διανυσματικές μηχανές υποστήριξης

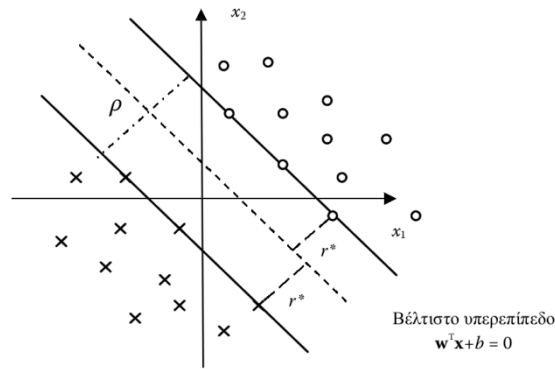
Οι διανυσματικές μηχανές υποστήριξης (SVMs), συμπεριλαμβανομένου του διανυσματικού κατηγοριοποιητή υποστήριξης (SVC) και του διανυσματικού παλινδρομητή υποστήριξης (SVR) ο οποίος χρησιμοποιείται σε μη γραμμικά προβλήματα παλινδρόμησης και δεν έγινε χρήση του στην παρούσα εργασία όποτε δεν θα επεκταθούμε σε αυτόν, είναι μεταξύ των πιο στιβαρών και ακριβών μεθόδων ανάμεσα σε όλους τους γνωστούς αλγορίθμους εξόρυξης δεδομένων. Οι SVM που αναπτύχθηκαν αρχικά από τον Vapnik στη δεκαετία του '90 [29-36], στηρίζονται σε ένα καλό θεωρητικό υπόβαθρο πάνω στη στατιστική θεωρία εκμάθησης, απαιτούν μόνο λίγα παραδείγματα για τους κανόνες εκμάθησης, και είναι συχνά ανεπηρέαστοι στον αριθμό των διαστάσεων. Στην προηγούμενη δεκαετία, οι SVM έχουν αναπτυχθεί με γρήγορο ρυθμό και θεωρητικά αλλά έχουν χρησιμοποιηθεί και σε διάφορες εφαρμογές.

### Διανυσματικός κατηγοριοποιητής υποστήριξης (SVC)

Για την εκμάθηση δύο κατηγοριών που είναι γραμμικά διαχωριζόμενες, ο στόχος του SVC είναι να βρεθεί ένα υπερεπίπεδο (hyperplane) το οποίο μπορεί να χωρίσει τις δύο κατηγορίες δεδομένων δειγμάτων με ένα μέγιστο περιθώριο που είναι αποδειγμένα ικανό να προσφέρει την καλύτερη δυνατότητα γενίκευσης. Η δυνατότητα γενίκευσης αναφέρεται στο γεγονός ότι ένας κατηγοριοποιητής όχι μόνο έχει καλή εκτέλεση κατηγοριοποίησης (π.χ., ακρίβεια) στα στοιχεία εκπαίδευσης, αλλά ακόμα εγγυάται και την υψηλή ακρίβεια πρόβλεψης για τα μελλοντικά δεδομένα από την ίδια κατανομή με τα στοιχεία εκπαίδευσης.

Διαισθητικά, ένα περιθώριο μπορεί να οριστεί ως το απαιτούμενο διάστημα, ή ο διαχωρισμός, μεταξύ των δύο κατηγοριών όπως καθορίζεται από το υπερεπίπεδο. Γεωμετρικά, το περιθώριο αντιστοιχεί στην ελάχιστη απόσταση μεταξύ των πιο κοντινών σημείων στοιχείων σε οποιοδήποτε σημείο του υπερεπιπέδου. Η Εικόνα 9 επεξηγεί με μια γεωμετρική κατασκευή το αντίστοιχο βέλτιστο υπερεπίπεδο υπό τις προϋποθέσεις που προαναφέραμε για ένα δισδιάστατο χώρο εισαγωγής.





Εικόνα 9 Αναπαράσταση του βέλτιστου υπερεπίπεδου SVC για την περίπτωση γραμμικά διαχωριζόμενου χώρου. (Πηγή <http://slideplayer.gr>)

Έστω  $w$  και  $b$  να ορίζουν αντίστοιχα το διάνυσμα στάθμης και την μεροληψία στο βέλτιστο υπερεπίπεδο. Το αντίστοιχο υπερεπίπεδο μπορεί να οριστεί ως:

$$w^T x + b = 0 \quad (1)$$

Η επιθυμητή προσανατολισμένη γεωμετρική απόσταση από το δείγμα  $x$  στο βέλτιστο υπερεπίπεδο [37] είναι :

$$r = \frac{g(x)}{\|w\|} \quad (2)$$

όπου  $g(x) = w^T x + b$  είναι η συνάρτηση διάκρισης [33] όπως καθορίζεται από το υπερεπίπεδο που επίσης καλείται και λειτουργικό περιθώριο του  $x$  δοθέντος των  $w$  και  $b$ . Συνεπώς, ο SVC στοχεύει να βρει τις παραμέτρους  $w$  και  $b$  για βέλτιστο υπερεπίπεδο προκειμένου να μεγιστοποιηθεί το περιθώριο του διαχωρισμού [ $\rho$  στην εξίσωση (5)] που καθορίζεται από τις πιο κοντινές γεωμετρικές αποστάσεις  $r^*$  από τις δύο κατηγορίες, αντίστοιχα, κατά συνέπεια ο SVC ονομάζεται και ως κατηγοριοποιητής μέγιστου περιθωρίου. Τώρα χωρίς βλάβη της γενικότητας, καθορίζουμε λειτουργικό περιθώριο [10] να είναι ίσο με 1 δηλαδή δεδομένου ενός συνόλου εκπαίδευσης  $\{x_i, y_i\}_{i=1}^n \in R^m \times \{\pm 1\}$  έχουμε:

$$\begin{cases} w^T x_i + b \geq 1, \text{ για } y_i = 1 \\ w^T x_i + b \leq -1, \text{ για } y_i = -1 \end{cases} \quad (3)$$

Τα συγκεκριμένα σημεία δεδομένων  $(x_i, y_i)$  για τα οποία ικανοποιούνται οι ισότητες των πρώτων ή δεύτερων μελών στις σχέσεις (3) καλούνται διανύσματα υποστήριξης, τα οποία είναι ακριβώς τα πιο κοντινά σημεία δεδομένων ως προς το βέλτιστο υπερεπίπεδο [11]. Κατόπιν, η αντίστοιχη γεωμετρική απόσταση από το διάνυσμα υποστήριξης  $x^*$  προς το βέλτιστο υπερεπίπεδο είναι :

$$r^* = \frac{g(\mathbf{x}^*)}{\|\mathbf{w}\|} = \begin{cases} \frac{1}{\|\mathbf{w}\|}, & \text{αν } y^* = 1 \\ -\frac{1}{\|\mathbf{w}\|}, & \text{αν } y^* = -1 \end{cases} \quad (4)$$

Στο σχήμα 3.1, το περιθώριο του διαχωρισμού  $\rho$  είναι:

$$\rho = 2r^* = \frac{2}{\|\mathbf{w}\|} \quad (5)$$

Για να εξασφαλιστεί ότι θα βρεθεί το μέγιστο εύρος του υπερεπίπεδου, ο SVC προσπαθεί να μεγιστοποιήσει το  $\rho$  σε σχέση με το  $\mathbf{w}$  και το  $b$ :

$$\max_{\mathbf{w}, b} \frac{2}{\|\mathbf{w}\|}$$

$$y_i(\mathbf{w}^T \mathbf{x}_i + b) \geq 1, \quad i = 1, 2, \dots, n \quad (6)$$

Ισοδύναμα:

$$\max_{\mathbf{w}, b} \frac{1}{2} \|\mathbf{w}\|^2$$

$$y_i(\mathbf{w}^T \mathbf{x}_i + b) \geq 1, \quad i = 1, 2, \dots, n \quad (7)$$

Γενικά, χρησιμοποιείτε συχνά η τετραγωνική νόρμα  $\|\mathbf{x}\|^2$  αντί της απλής  $\|\mathbf{x}\|$  για λόγους ευκολίας για την πραγματοποίηση των επόμενων βημάτων βελτιστοποίησης.

Γενικά, λύνουμε το περιορισμένο πρόβλημα βελτιστοποίησης στην εξίσωση (7), γνωστό και ως πρωταρχικό πρόβλημα, με τη χρησιμοποίηση της μεθόδου των πολλαπλασιαστών Lagrange. Κατασκευάζουμε έτσι την ακόλουθη συνάρτηση Lagrange:

$$L(\mathbf{w}, b, \mathbf{a}) = \frac{1}{2} \mathbf{w}^T \mathbf{w} - \sum_{i=1}^n a_i [(\mathbf{w}^T \mathbf{x}_i + b) - 1] \quad (8)$$

όπου  $a_i$  είναι ο πολλαπλασιαστής Lagrange της  $i$ -ανισότητας.

Παραγωγίζοντας την  $L(\mathbf{w}, b, \mathbf{a})$  ως προς  $\mathbf{w}$  και  $b$ , και θέτοντας τα αποτελέσματα ίσα με μηδέν, παίρνουμε τις ακόλουθες δύο συνθήκες βελτιστοποίησης:

$$\begin{cases} \frac{\partial L(\mathbf{w}, b, \mathbf{a})}{\partial \mathbf{w}} = 0 \\ \frac{\partial L(\mathbf{w}, b, \mathbf{a})}{\partial b} = 0 \end{cases} \quad (9)$$

Τότε παίρνουμε:

$$\begin{cases} \mathbf{w} = \sum_{i=1}^n a_i y_i \mathbf{x}_i \\ \sum_{i=1}^n a_i y_i = 0 \end{cases} \quad (10)$$

Αντικαθιστώντας την εξίσωση (10) στη συνάρτηση Lagrange (8), μπορούμε να πάρουμε το αντίστοιχο δυϊκό πρόβλημα:

$$\max_{\mathbf{w}, b} W(a) = \sum_{i=1}^n a_i - \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n a_i a_j y_i y_j \mathbf{x}_i^T \mathbf{x}_j$$

$$\sum_{i=1}^n a_i y_i = 0$$

$$a_i \geq 0, \quad i = 1, 2, \dots, n \quad (11)$$

Και ταυτόχρονα, η συμπληρωματική συνθήκη Karush-Kuhn-Tucker είναι:

$$a_i [(\mathbf{w}^T \mathbf{x}_i + b) - 1] = 0, \quad i = 1, 2, \dots, n \quad (12)$$

Συνεπώς, μόνο τα διανύσματα στήριξης ( $\mathbf{x}_i, y_i$ ) που είναι τα πιο κοντινά δεδομένα στο βέλτιστο υπερεπίπεδο και καθορίζουν το μέγιστο εύρος, αντιστοιχούν στα μη μηδενικά  $a_i$ 's. Όλος ο υπόλοιπα  $a_i$ 's είναι ίσα με μηδέν. Το δυϊκό πρόβλημα στην εξίσωση (11) είναι ένα χαρακτηριστικό πρόβλημα βελτιστοποίησης κυρτού τετραγωνικού προγραμματισμού. Σε πολλές περιπτώσεις, μπορεί να συγκλίνει αποτελεσματικά στο ολικό βέλτιστο με την υιοθέτηση μερικών κατάλληλων τεχνικών βελτιστοποίησης, όπως ο αλγόριθμος βελτιστοποίησης διαδοχικών ελαχίστων(SMO) [33].

Μετά αφού προσδιοριστούν οι βέλτιστοι πολλαπλασιαστές Lagrange  $a_i^*$ , μπορούμε να υπολογίσουμε το βέλτιστο διάνυσμα στάθμης  $\mathbf{w}^*$  από την εξίσωση (10):

$$\mathbf{w}^* = \sum_{i=1}^n a_i^* y_i \mathbf{x}_i \quad (13)$$

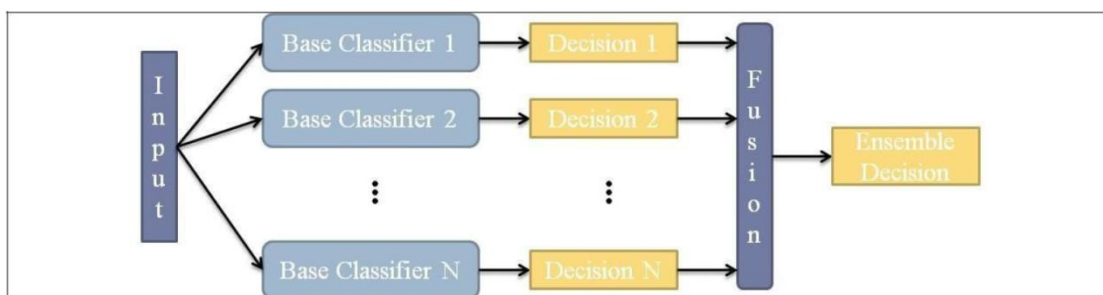
Κατόπιν, χρησιμοποιώντας ένα θετικό διάνυσμα στήριξης  $\mathbf{x}_s$ , η αντίστοιχη βέλτιστη μεροληψία δίνεται από τη σχέση:

$$b^* = 1 - \mathbf{w}^{*T} \mathbf{x}_s, \quad \text{για } y_s = 1 \quad (14)$$

## ΚΑΤΗΓΟΡΙΟΠΟΙΗΤΕΣ ΣΥΝΟΛΩΝ

Ο κατηγοριοποιητής συνόλων αναφέρεται σε μια ομάδα μεμονωμένων κατηγοριοποιητών που εκπαιδεύονται συνεργατικά ό ένας με τον άλλο στο σύνολο στοιχείων σε ένα εποπτευμένο(supervised) πρόβλημα κατηγοριοποίησης. Εδώ θα παρουσιάσουμε μια ανασκόπηση των πιο συνηθισμένων κατηγοριοποιητών συνόλων στη διεθνή βιβλιογραφία. Μερικοί κατηγοριοποιητές συνόλων αναπτύσσονται επίσης απευθυνόμενοι σε συγκεκριμένες εφαρμογές.

Ένα εποπτευμένο πρόβλημα ταξινόμησης εμπίπτει στην κατηγορία εκμάθησης από περιπτώσεις όπου κάθε περίπτωση-πρότυπο συνδέεται με μια κατηγορία. Συμβατικά ένας μεμονωμένος ταξινομητής, όπως το νευρωνικό δίκτυο, το δέντρο απόφασης, ή μια διανυσματική μηχανή υποστήριξης εκπαιδεύεται σε ένα σύνολο στοιχείων. Ανάλογα με τη κατανομή των μοτίβων που παρουσιάζονται, είναι πιθανό ότι δεν μαθαίνονται καλά όλα τα πρότυπα από έναν μεμονωμένο κατηγοριοποιητή. Ο κατηγοριοποιητής δεν αποδίδει καλά στη δοκιμή που του ανατίθεται κάτω από τέτοια σενάρια. Μια λύση σε αυτό το πρόβλημα είναι να εκπαιδευθεί μια ομάδα κατηγοριοποιητών στο ίδιο πρόβλημα. Η υπάρχουσα βιβλιογραφία χρησιμοποιεί τον όρο «κατηγοριοποιητής συνόλων» για να αναφερθεί σε ομάδα κατηγοριοποιητών [12]. Οι μεμονωμένοι κατηγοριοποιητές καλούνται κατηγοριοποιητές βάσης ή ασθενείς κατηγοριοποιητές. Κατά τη διάρκεια της εκμάθησης, οι κατηγοριοποιητές βάσης εκπαιδεύονται χωριστά στο σύνολο στοιχείων. Κατά τη διάρκεια της πρόβλεψης, οι κατηγοριοποιητές βάσης παρέχουν μια απόφαση σχετικά με ένα δοκιμαστικό μοτίβο. Μια συνδυαστική μέθοδος αναλύει μετά τις αποφάσεις που παράγονται από τους εκάστοτε κατηγοριοποιητές βάσεων. Υπάρχει ένας μεγάλος αριθμός συνδυαστικών μεθόδων στη βιβλιογραφία συμπεριλαμβανομένης της ψηφοφορίας με πλειοψηφία, της αρίθμησης Borda, των αλγεβρικών συνδυαστών κ.λ.π. [12].



Εικόνα 10 Μηχανισμός λειτουργίας κατηγοριοποιητών συνόλων (Πηγή <http://slideplayer.gr>)

Η φιλοσοφία του κατηγοριοποιητή συνόλων είναι ότι ένας κατηγοριοποιητής βάσης αντισταθμίζει τα λάθη που γίνονται από έναν διαφορετικό κατηγοριοποιητή βάσης. Ωστόσο η εκπαίδευση του κατηγοριοποιητή βάσης κατά τρόπο απλό δεν πρόκειται να λύσει αυτό το πρόβλημα. Όπως επισημαίνεται [12], ένας κατηγοριοποιητής συνόλου αποδίδει καλύτερα από τους αντίστοιχους των βάσεων που αποτελείται ακόμα και εάν οι ταξινομητές βάσεων είναι ακριβείς και διαφορετικοί σε δομή. Η ποικιλομορφία αναφέρεται στο γεγονός ότι τα λάθη κατηγοριοποιητών βάσης είναι ασύνδετα. Υπάρχουν διάφοροι τρόποι για να υπολογιστεί η ποικιλομορφία συμπεριλαμβανομένων των μέτρων ποικιλομορφίας για ζεύγη (π.χ. Q-statistics, ο συντελεστής συσχέτισης, το διπλό μέτρο ελαττωμάτων) άλλα και των μέτρων ποικιλομορφίας για μη-ζευγή (π.χ. το μέτρο εντροπίας, η διακύμανση Kohavi-Wolpert, κ.α.) [13,14]. Οι διαφορετικές μέθοδοι παραγωγής κατηγοριοποιητών συνόλου στοχεύουν να επιτύχουν την ποικιλομορφία μεταξύ των κατηγοριοποιητών βάσης. Μερικοί κατηγοριοποιητές συνόλων αναπτύσσονται απευθυνόμενοι για συγκεκριμένα προβλήματα/εφαρμογές. Ακολουθώς παρουσιάζουμε συνοπτικά τους διαφορετικούς κατηγοριοποιητές βάσεων όπως τους κατηγοριοποιητές συνόλων.

#### Κατηγοριοποιητές Βάσης (Base Classifiers)

Οι κατηγοριοποιητές βάσεων αναφέρονται στους μεμονωμένους κατηγοριοποιητές που χρησιμοποιούνται για να κατασκευάσουν τους κατηγοριοποιητές συνόλων. Το νευρωνικό δίκτυο, η διανυσματική μηχανή υποστήριξης, και οι ταξινομητές k-NN είναι μερικοί από τους συνήθως χρησιμοποιημένους κατηγοριοποιητές βάσεων. Για λόγους πληρότητας εξηγούμε εν συντομία τη διαδικασία εκπαίδευσης και δοκιμής αυτών των κατηγοριοποιητών βάσεων. Στην ταξινόμηση k-NN υπολογίζεται η απόσταση μεταξύ ενός δοκιμαστικού μοτίβου και όλων των μοτίβων που εμφανίζονται στο σύνολο εκπαίδευσης. Η απόσταση μπορεί να υπολογιστεί χρησιμοποιώντας την ευκλείδεια απόσταση ή την απόσταση Manhattan. Οι πιθανές κατηγορίες λαμβάνουν μια ψήφο από κάθε ένα από τα K-μοτιβα που είναι τα πιο κοντινά στο δοκιμαστικό από άποψη απόστασης. Η κατηγορία που λαμβάνει την υψηλότερη ψηφοφορία θεωρείται η κατηγορία του προτύπου δοκιμής.

Ένα νευρωνικό δίκτυο [15] μπορεί να θεωρηθεί ως σύστημα υπολογισμού φτιαγμένο από διάφορα απλά, ιδιαίτερα διασυνδεδεμένα και επεξεργασμένα στοιχεία, το οποίο επεξεργάζεται τις πληροφορίες από την αντίδραση στη δυναμική τους κατάσταση σε εξωτερικά ερεθίσματα. Τα νευρωνικά δίκτυα έχουν στρωματοποιημένη δομή. Τα επίπεδα αποτελούνται από διάφορους διασυνδεδεμένους κόμβους που περιέχουν μια λειτουργία ενεργοποίησης. Τα πρότυπα παρουσιάζονται στο δίκτυο μέσω του επιπέδου εισαγωγής, το οποίο τα

διαβιβάζει σε ένα ή περισσότερα κρυμμένα επίπεδα όπου η πραγματική επεξεργασία γίνεται μέσω ενός συστήματος σταθμισμένων συνδέσεων. Τα κρυμμένα επίπεδα συνδέονται μετά με ένα επίπεδο εξόδου όπου η δίνεται η τελική απάντηση. Τα περισσότερα νευρωνικά δίκτυα περιέχουν έναν κανόνα εκμάθησης που τροποποιεί τη συσχέτιση των συνδέσεων σύμφωνα με τα πρότυπα εισόδου που τα συνοδεύουν.

Ένα SVM [16] ταξινομεί τα στοιχεία με το μετασχηματισμό των στοιχείων σε υψηλότερη διάσταση χρησιμοποιώντας μια συνάρτηση πυρήνα και βρίσκοντας έπειτα καλύτερο υπερεπίπεδο που διαχωρίζει τα πρότυπα μιας κατηγορίας από εκείνα μιας άλλης κατηγορίας. Καλύτερο υπερεπίπεδο για ένα SVM –όπως έχουμε προαναφέρει - είναι αυτό με το μέγιστο περιθώριο μεταξύ των κατηγοριών. Το περιθώριο σημαίνει το μέγιστο πλάτος που ορίζουν οι παράλληλες ευθείες του υπερεπιπέδου που δεν έχει περιέχει εσωτερικά κανένα πρότυπο. Τα διανύσματα στήριξης είναι τα σημεία των δεδομένων που είναι τα πιο κοντινά στο όριο των ευθειών που ορίζουν το υπερεπίπεδο

#### Κατηγοριοποιητές Συνόλων (Ensemble Classifiers)

Η χρησιμοποίηση μεθόδων παραγωγής για κατηγοριοποιητές συνόλων μπορεί να ταξινομηθεί γενικά σε έξι ομάδες [17] που είναι βασισμένες (i) στο χειρισμό των παραμέτρων εκπαίδευσης των κατηγοριοποιητών, (ii) στον χειρισμό της λειτουργίας του σφάλματος, (iii) στο χειρισμό του διαστήματος των χαρακτηριστικών γνωρισμάτων, (iv) στο χειρισμό του χαρακτηρισμού των δεδομένων εξόδου, (v) στην ανάλυση των συστάδων, και τέλος στον (vi) στο χειρισμό των προτύπων εκπαίδευσης.

Παραγωγή κατηγοριοποιητών συνόλων από το χειρισμό των παραμέτρων εκπαίδευσης

Η ποικιλομορφία μπορεί να επιτευχθεί με το χειρισμό των παραμέτρων εκπαίδευσης των ταξινομητών βάσης σε ένα σύνολο. Τα διαφορετικά χαρακτηριστικά των δικτύων χρησιμοποιούνται για να εκπαιδεύσουν διαδικασία εκμάθησης των βασικών νευρωνικών δικτύων μια διαδικασία που περιγράφεται στις εργασίες [18] και [19]. Αυτές οι μέθοδοι επιτυγχάνουν καλύτερη γενίκευση.

Παραγωγή κατηγοριοποιητών συνόλων από το χειρισμό της λειτουργίας σφαλμάτων

Μια ομάδα μεθόδων κατασκευής κατηγοριοποιητών συνόλων αντιμετωπίζει αυτό το ζήτημα με την αύξηση της λειτουργίας σφάλματος των κατηγοριοποιητών βάσεων. Ένα σφάλμα επιβάλλεται εάν οι κατηγοριοποιητές βάσεων κάνουν τα ίδια

λάθη σε παρόμοια πρότυπα. Ο αρνητικός συσχετισμός στη διαδικασία της εκπαίδευσης ([20] , [21]) είναι ένα τέτοιο σύνολο όπου όλα τα μεμονωμένα δίκτυα στο σύνολο εκπαιδεύονται ταυτόχρονα και αμφίδρομα μέσω των όρων του περιορισμού συσχέτισης στις συναρτήσεις των σφαλμάτων τους.

Παραγωγή κατηγοριοποιητών συνόλων από το χειρισμό του διαστήματος χαρακτηριστικών γνωρισμάτων

Σε μια άλλη ομάδα κατηγοριοποιητών συνόλων η ποικιλομορφία μεταξύ των ταξινομητών βάσεων επιτυγχάνεται με το χειρισμό του διαστήματος χαρακτηριστικών γνωρισμάτων εισόδου. Τα διαφορετικά υποσύνολα των χαρακτηριστικών γνωρισμάτων χρησιμοποιούνται για να εκπαιδεύσουν τους κατηγοριοποιητές βάσεων όπως μπορεί να βρει κάποιος στα [22-24]. Οι τυχαίοι υποχώροι των κατηγοριοποιητών συνόλων έχουν γενικά χαμηλότερη απόδοση από ότι άλλοι κατηγοριοποιητές συνόλων.

Παραγωγή κατηγοριοποιητών συνόλων με χειρισμό του χαρακτηρισμού των δεδομένων εξόδου

Οι ταξινομητές συνόλων μπορούν να κατασκευαστούν από το χειρισμό στοχευμένων δεδομένων εξόδου όπως αναφέρεται στις εργασίες [25] και [26]. Στα σύνολα εναλλαγής χαρακτηρισμών [25], κάθε κατηγοριοποιητής βάσεων παράγεται με την αλλαγή του χαρακτηρισμού της κατηγορίας ενός μέρους των προτύπων εκπαίδευσης που επιλέγονται τυχαία από το αρχικό σύνολο εκπαίδευσης.

Παραγωγή κατηγοριοποιητών συνόλων με συσταδοποίηση

Οι κατηγοριοποιητές συνόλων μπορούν να παραχθούν ακόμα με τον διαμερισμό του συνόλου εκπαίδευσης σε μη-επικαλύπτομενες συστάδες και εκπαίδευση των κατηγοριοποιητών βάσεων πάνω σε αυτές. Αυτοί οι ταξινομητές καλούνται συσταδοποιημένα σύνολα [27-31]. Τα πρότυπα που τείνουν να μείνουν κοντά στον ευκλείδιο χώρο προσδιορίζονται με αυτήν την διαδικασία με φυσικό τρόπο εκ κατασκευής της. Ένα πρότυπο μπορεί να ανήκει σε μια συστάδα μόνο που μια επιλογή προσέγγισης ακολουθείται για τη λήψη της απόφασης της κατηγορίας των συνόλων. Αυτές οι μέθοδοι στοχεύουν να μειώσουν την πολυπλοκότητα εκμάθησης των μεγάλων συνόλων με πολλά στοιχεία [27]. Τα συσταδοποιημένα σύνολα [28]-[31] δεν παρέχουν οποιοδήποτε μηχανισμό για το βέλτιστο αριθμό συστάδων.

Μερικοί ερευνητές [32-38] παρέχουν έναν μηχανισμό για να εκτελεστεί ένας ομαλός διαχωρισμός του συνόλου στοιχείων που οδηγεί στην καλύτερη απόδοση ταξινόμησης [32]. Οι κατηγοριοποιητές συνόλων στις εργασίες [33-38] όπως μπορούμε να δούμε παράγονται από (i) διαμέριση των στοιχείων σε συστάδες σε διαφορετικά επίπεδα, (ii) εκπαιδεύοντας κατηγοριοποιητές βάσεων στις συστάδες στα διαφορετικά επίπεδα.

Κατά τη διάρκεια της πρόβλεψης (I) η κοντινότερη συστάδα σε κάθε επίπεδο βρίσκεται για το δοκιμαστικό πρότυπο, (II) η πρόβλεψη λήφθηκε από τους κατηγοριοποιητές στις κοντινότερες συστάδες σε κάθε επίπεδο, και (III) οι αποφάσεις από τα επίπεδα συγχωνεύονται σε μια ενιαία απόφαση χρησιμοποιώντας την ψηφοφορία με πλειοψηφία. Ακόμα στις [33-35] έχουμε τους ίδιους αριθμούς συστάδων να χρησιμοποιούνται στα διαφορετικά επίπεδα ενώ στην [38] έγινε χρήση και διαφορετικού πλήθους συστάδων. Τέλος η βελτιστοποίηση του αριθμού συστάδων και τα επίπεδα εξετάζονται στις εργασίες [36-38].

Παραγωγή κατηγοριοποιητών συνόλων από το χειρισμό των προτύπων εκπαίδευσης

Η μεγαλύτερη κατηγορία-μεθοδολογία παραγωγής κατηγοριοποιητών συνόλων είναι αυτή με το χειρισμό των προτύπων εκπαίδευσης όπου οι κατηγοριοποιητές βάσεων εκπαιδεύονται στα διαφορετικά υποσύνολα των προτύπων εκπαίδευσης. Οι μέθοδοι διαφέρουν στην παραγωγή των υποσυνόλων. Στη μέθοδο Bootstrap Aggregating (BAGGING) [39] τα υποσύνολα εκπαίδευσης προέρχονται τυχαία (με την αντικατάσταση) από το γενικευμένο σύνολο εκπαίδευσης. Οι ομοιογενείς κατηγοριοποιητές βάσεων εκπαιδεύονται στα εκάστοτε υποσύνολα. Υπάρχουν διάφορες παραλλαγές των προσεγγίσεων BAGGING και AGGRAGATING συμπεριλαμβανομένων των τυχαίων δασών [40] και της τοποθέτησης BAGGING μεγάλης κλίμακας [41]. Η τεχνική BOOSTING [42] δημιουργεί τα υποσύνολα δεδομένων για την εκπαίδευση κατηγοριοποιητών βάσεων με τη λήψη δείγματος από τα πρότυπα εκπαίδευσης, παρέχοντας το πρότυπο εκπαίδευσης που περιέχει της περισσότερες πληροφορίες για κάθε διαδοχικό ταξινομητή. Σε κάθε ένα από τα πρότυπα εκπαίδευσης ορίζεται ένα μέτρο που καθορίζει πόσο καλά η περίπτωση ταξινομήθηκε στην προηγούμενη επανάληψη. Το δεδομένο εκπαίδευσης που είναι λανθασμένα ταξινομημένο συμπεριλαμβάνεται στο υποσύνολο εκπαίδευσης για την επόμενη επανάληψη. Τέλος υπάρχει και η τεχνική AdaBOOST [43] που είναι μια γενικευμένη έκδοση της BOOSTING.



## Κεφάλαιο 5.3 Εκπαίδευση αλγορίθμου εξόρυξης δεδομένων για ατυχήματα LNG-LPG πλοίων

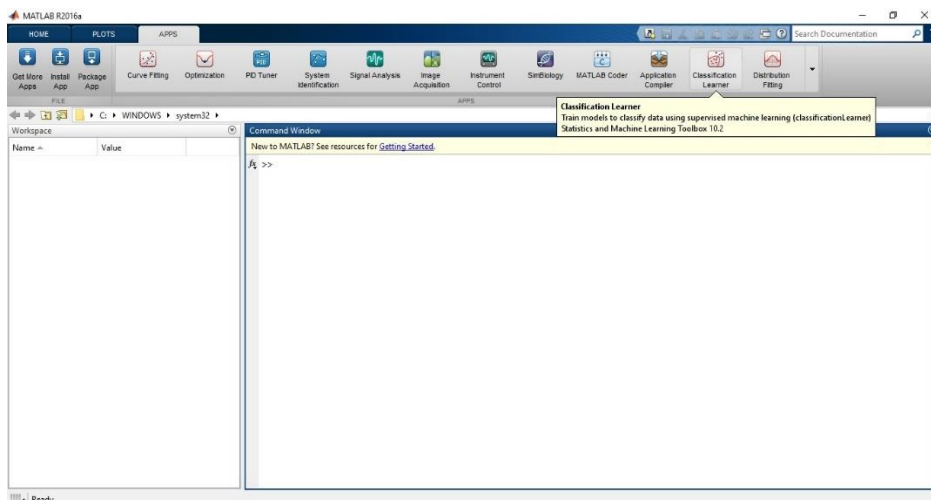
Σε αυτό το κεφάλαιο θα γίνει μία προσπάθεια να εφαρμόσουμε τις τεχνικές της εξόρυξης δεδομένων που αναλύθηκαν στα ατυχήματα που αφορούν την παρούσα διπλωματική. Συγκεκριμένα θα γίνει μία προσπάθεια εκπαίδευσης ενός αλγορίθμου ο οποίος θα μας υπολογίζει με την μέγιστη δυνατή ακρίβεια τον τύπο ατυχήματος που είναι πιο πιθανό να συμβεί σε ένα πλοίο με συγκεκριμένα χαρακτηριστικά. Τα χαρακτηριστικά τα οποία θα χρησιμοποιηθούν για την εκπαίδευση του αλγορίθμου είναι τα ακόλουθα.

- Ηλικία του πλοίου όταν συνέβη το ατύχημα (Age of Ship)
- Χώρα κατασκευής του πλοίου (Country Built)
- Ζώνη που συνέβη το ατύχημα (Zone of Accident)
- Είδος του πλοίου (Type of Ship)
- Σημαία του πλοίου (Flag of Ship)

Τα χαρακτηριστικά αυτά είναι τα ίδια που αποδείξαμε στο κεφάλαιο 3 ότι επηρεάζουν τον τύπο του ατυχήματος χρησιμοποιώντας την μέθοδο  $\chi^2$ . Για την εφαρμογή οποιασδήποτε μεθόδου εξόρυξης δεδομένων είναι αναγκαία η απόδειξη ότι τα αίτια (observers) επηρεάζουν τα αποτελέσματα (responses). Για την εκπαίδευση του αλγορίθμου χρησιμοποιήθηκε το λογισμικό Matlab R2016a. Η παρούσα έκδοση της Matlab και όλες οι ακόλουθες έχουν ενσωματωμένη εφαρμογή για την εκπαίδευση αλγορίθμων με σκοπό την εξόρυξη δεδομένων. Συγκεκριμένα δίνεται η δυνατότητα στον χρήστη να εκπαιδεύσει ένα πλήθος αλγορίθμων γρήγορα και εύκολα χρησιμοποιώντας έναν και μόνο πίνακα εισαγωγής ενώ ταυτόχρονα του γνωστοποιείται ποιος από τους αλγόριθμους έχει την μεγαλύτερη ακρίβεια (Accuracy) για τα δεδομένα τα οποία παρείχε. Θα ακολουθήσει ένα παράδειγμα εκπαίδευσης αλγορίθμων με χρήση Matlab και στην συνέχεια θα παρουσιαστούν τα αποτελέσματα που προέκυψαν για διάφορες περιπτώσεις εισαγωγής δεδομένων.

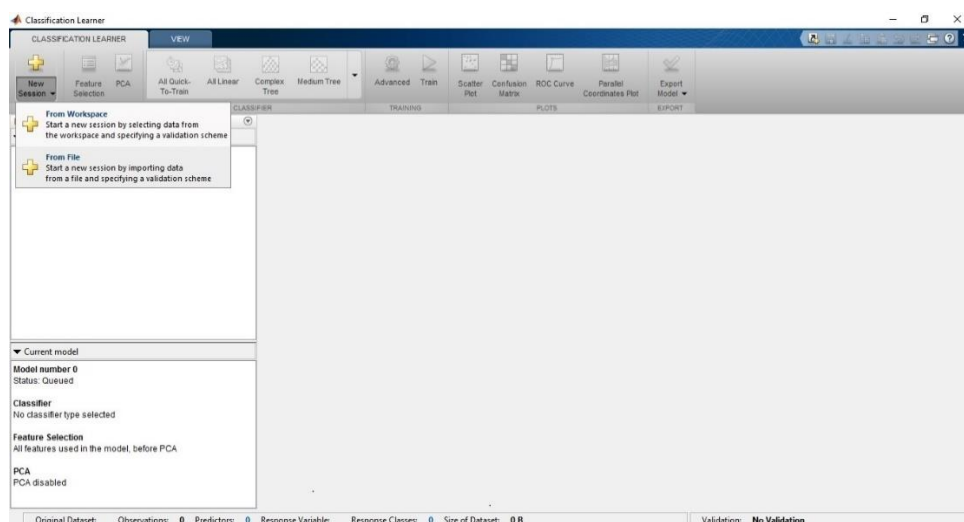
### 5.3.1 Μέθοδος εκπαίδευσης και επιλογής αλγορίθμου

Στις ακόλουθες εικόνες φαίνεται σταδιακά η μεθοδολογία εκπαίδευσης και επιλογής αλγορίθμων με χρήση του προγράμματος Matlab.



Εικόνα 11 Πρώτο βήμα εκπαίδευσης αλγορίθμου

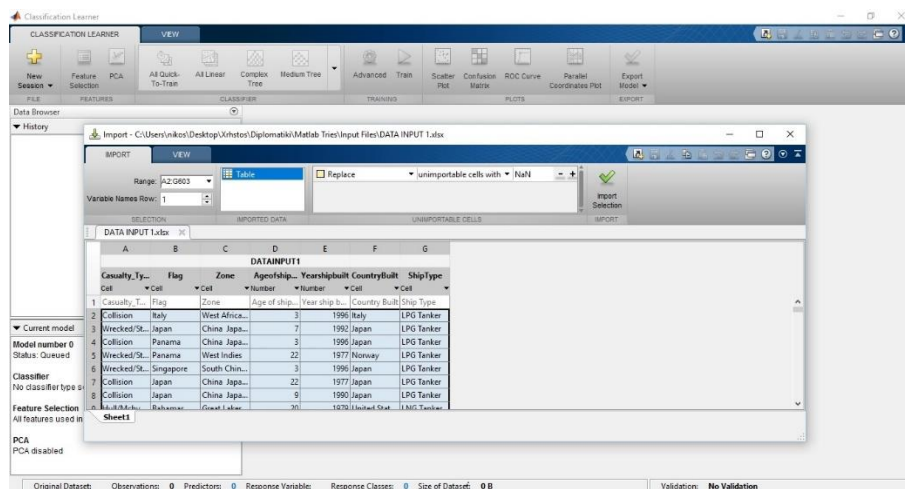
Αρχικά ανοίγουμε το πρόγραμμα Matlab. Στο τρίτο Tab του προγράμματος (APPS) διαλέγουμε την επιλογή “Classification Learner” όπως φαίνεται στην Εικόνα 11.



Εικόνα 12 Δεύτερο βήμα εκπαίδευσης αλγορίθμου

## Στατιστική Ανάλυση Ναυτικών Ατυχημάτων σε πλοία LNG-LPG και Ανάπτυξη Αλγορίθμου Εξόρυξης Δεδομένων

Στην συνέχεια εισάγουμε το αρχείο excel το οποίο περιέχει τα δεδομένα εκπαίδευσης του αλγορίθμου όπως φαίνεται στην Εικόνα 12.



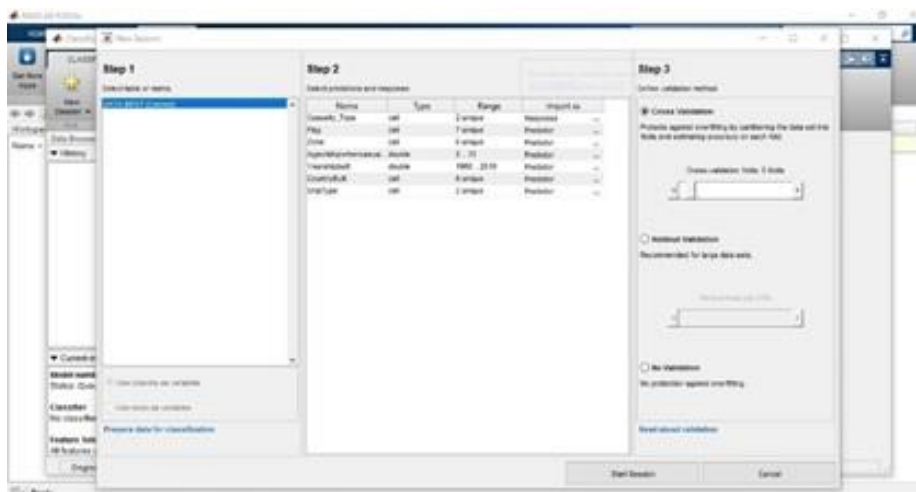
Εικόνα 13 Τρίτο βήμα εκπαίδευσης αλγορίθμου

Το αρχείο excel το οποίο θα εισαχθεί έχει την ακόλουθη μορφολογία όπως φαίνεται και στην Εικόνα 13: Η κάθε γραμμή του αρχείου αντιστοιχεί σε ένα ατύχημα. Η πρώτη στήλη του αρχείου εμπεριέχει για το κάθε ατύχημα το αποτέλεσμα (response) ενώ οι υπόλοιπες στήλες εμπεριέχουν τα αίτια(observers) που οδήγησαν στο αποτέλεσμα.

Casualty_Type	Flag	Zone	Ageofship	Yearshpbuilt	CountryBuilt	ShipType
Collision	Italy	West Africa...	3	1996	Italy	LPG Tanker
Wrecked/St...	Japan	China Japa...	7	1992	Japan	LPG Tanker
Collision	Panama	China Japa...	3	1996	Japan	LPG Tanker
Wrecked/St...	Panama	West Indies	22	1977	Norway	LPG Tanker
Wrecked/St...	Singapore	South Chin...	3	1996	Japan	LPG Tanker
Collision	Japan	China Japa...	22	1977	Japan	LPG Tanker
Collision	Japan	China Japa...	9	1990	Japan	LPG Tanker
Hull/Mchy...	Bahamas	Great Lake...	20	1979	United Stat...	LPG Tanker
Hull/Mchy...	Japan	China Japa...	15	1985	Japan	LPG Tanker
Collision	Japan	China Japa...	9	1991	Japan	LPG Tanker
Fire/Explosi...	Japan	China Japa...	7	2000	Japan	LNG Tanker
Collision	Panama	China Japa...	5	1995	South Korea	LNG Tanker
Contact	Norway	Iceland	26	1974	Norway	LNG Tanker
Hull/Mchy...	Belice	China Japa...	21	1979	Japan	LPG Tanker
Collision	South Korea	China Japa...	26	1971	Japan	LPG Tanker
Collision	China Hon...	Brjides N.S...	2	1998	Japan	LPG Tanker
Hull/Mchy...	Norway (Nis)	Brjides N.S...	24	1976	Federal Re...	LPG Tanker
Wrecked/St...	Japan	China Japa...	3	1997	Japan	LPG Tanker
Hull/Mchy...	Algeria	Brjides N.S...	20	1981	France	LNG Tanker

Εικόνα 14 Τέταρτο βήμα εκπαίδευσης αλγορίθμου

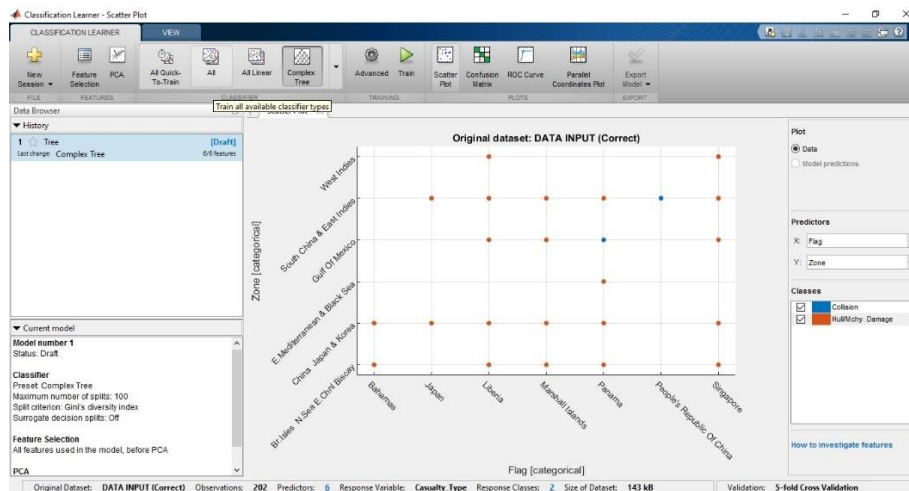
Σε αυτό το σημείο μας δίνεται η επιλογή να παραλείψουμε την εισαγωγή κάποιων στηλών ή και γραμμών. Στην δική μας περίπτωση αυτό δεν ήταν αναγκαίο καθώς τα excel που εισήχθησαν μορφοποιήθηκαν καταλλήλως πριν την εισαγωγή τους στο πρόγραμμα. Όταν τα δεδομένα μας έχουν την μορφή που επιθυμούμε τα εισάγουμε στο πρόγραμμα όπως φαίνεται στην Εικόνα 14.



Εικόνα 15 Πέμπτο βήμα εκπαίδευσης αλγορίθμου

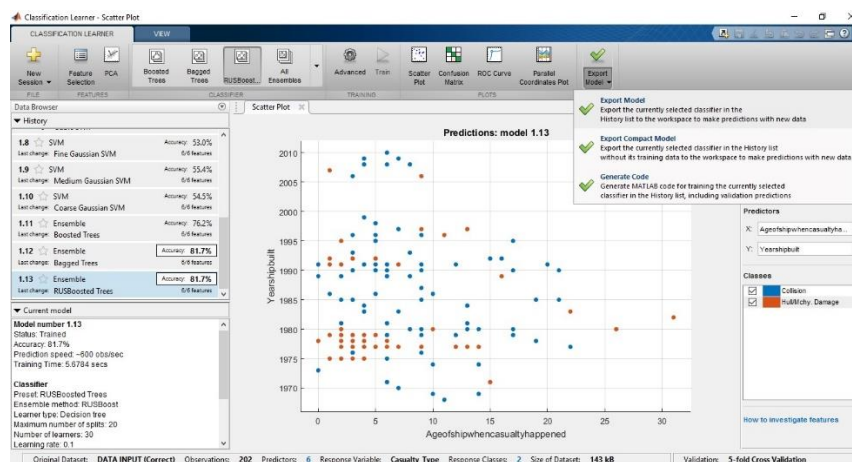
Σε αυτό το σημείο μας δίνεται η δυνατότητα να επιλέξουμε Validation method για την εκπαίδευση του αλγορίθμου μας. Μας δίνονται τρεις τρόποι Validation. Η πρώτη επιλογή είναι να μην έχουμε καθόλου Validation. Γενικά αυτή η επιλογή είναι ανεπιθύμητη καθώς καταλήγει σε ανακριβής αλγορίθμους με περιορισμένες δυνατότητες ακριβής πρόβλεψης. Η δεύτερη επιλογή ονομάζεται “Holdout Validation” και χρησιμοποιείται για μεγάλο όγκο δεδομένων όπου θα είχε πολύ μεγάλο υπολογιστικό κόστος η χρησιμοποίηση της τρίτης μεθόδου. Η Τρίτη και τελευταία μέθοδος η οποία χρησιμοποιήθηκε στα δεδομένα μας ονομάζεται “Cross Validation”. Σε αυτήν την μέθοδο χωρίζουμε τα δεδομένα σε υποσύνολα  $n$ . Στην συνέχεια ο αλγόριθμος εκπαιδεύεται για κάθε ένα πλήθος συνόλων εξαιρουμένου ενός από τα υποσύνολα. Το υποσύνολο που δεν χρησιμοποιήθηκε κάθε φορά για την εκπαίδευση του αλγορίθμου χρησιμοποιείται για τον υπολογισμό του Accuracy του. Δηλαδή γνωρίζοντας το αποτέλεσμα (response) του υποσυνόλου που δεν χρησιμοποιήθηκε και τα αντίστοιχα αίτια του (observers) υπολογίζουμε πόσα από τα δεδομένα μας θα προβλέπονταν επιτυχώς. Με την μέθοδο αυτή ο κάθε αλγόριθμος έχει την βέλτιστη ικανότητα πρόβλεψης αποτελεσμάτων και η εφαρμογή του Classification Learner μας επιστρέφει τις πιο ακριβείς τιμές Accuracy.

## Στατιστική Ανάλυση Ναυτικών Ατυχημάτων σε πλοία LNG-LPG και Ανάπτυξη Αλγορίθμου Εξόρυξης Δεδομένων



Εικόνα 16 Έκτο βήμα εκπαίδευσης αλγορίθμου

Στην συνέχεια μας δίνεται η δυνατότητα να επιλέξουμε ποιους από τους αλγορίθμους θέλουμε να εκπαιδύσουμε με τα δεδομένα που εισήγαμε. Ιδιαίτερα σημαντικό και εύχρηστο είναι το γεγονός ότι με την εισαγωγή ενός και μόνο πίνακα δεδομένων με χρήση Matlab μας δίνεται η δυνατότητα να εκπαιδύσουμε ένα πλήθος αλγορίθμων. Στην συγκεκριμένη περίπτωση επιλέξαμε να εκπαιδύσουμε όλους τους διαθέσιμους αλγορίθμους. Καθώς η ποσότητα των δεδομένων που έχουμε στην διάθεση μας είναι περιορισμένη η εκπαίδευση του κάθε αλγορίθμου δεν έχει μεγάλο υπολογιστικό κόστος.



Εικόνα 17 Έβδομο βήμα εκπαίδευσης αλγορίθμου

Μετά την εκπαίδευση των αλγορίθμων λαμβάνουμε την τιμή του Accuracy για κάθε έναν από τους αλγόριθμους. Ο αλγόριθμος με το μεγαλύτερο Accuracy είναι και ο βέλτιστος για την δική μας μελέτη. Τέλος εξάγουμε τον εκπαιδευμένο κώδικα του εν λόγω αλγορίθμου πίσω στο Matlab ώστε να τον χρησιμοποιήσουμε για την εξαγωγή συμπερασμάτων.

### 5.3.2 Εκπαίδευση και επιλογή αλγορίθμου για τα ατυχήματα LNG-LPG πλοίων

Σε αυτό το κεφάλαιο θα γίνει μία εφαρμογή της μεθόδου που αναλύθηκε στα προηγούμενα κεφάλαια. Στόχος είναι η εκπαίδευση ενός αλγορίθμου ο οποίος θα μπορεί με ικανοποιητική ακρίβεια να προβλέψει το είδος ατυχήματος το οποίο είναι πιο πιθανό να συμβεί σε κάποιο πλοίο δοθέντων ορισμένων δεδομένων. Συνολικά πραγματοποιήθηκαν 6 δοκιμές εκπαίδευσης ενός αλγορίθμου με 5 διαφορετικές ομάδες δεδομένων. Σε όλες τις απόπειρες ανάπτυξης ενός αλγορίθμου χρησιμοποιήθηκε η μέθοδος “Cross Validation” με 50 folds για την παραγωγή του μέγιστου δυνατού accuracy. Δηλαδή το πλήθος των δεδομένων χωρίστηκε σε 50 ομάδες και μελετήθηκε με την διαδικασία που περιεγράφηκε στο προηγούμενο κεφάλαιο. Σκοπός μας είναι με την χρήση αυτής της μεθόδου η εκπαίδευση ενός αλγορίθμου με ικανοποιητικό Accuracy. Όταν χρησιμοποιείται αυτή η μέθοδος το παραγόμενο Accuracy αποτελεί μία πολύ καλή προσέγγιση της ικανότητας του αλγορίθμου να προβλέψει με επιτυχία.

#### Δοκιμή Πρώτη

Αρχικά χρησιμοποιήθηκαν για την εκπαίδευση του αλγορίθμου όλα τα ατυχήματα τα οποία είχαμε στην διάθεση μας στα οποία υπήρχε πληρότητα των δεδομένων. Τα δεδομένα τα οποία χρησιμοποιήθηκαν για την εκπαίδευση του αλγορίθμου είναι τα ίδια που χρησιμοποιήθηκαν και στο κεφάλαιο 4 της παρούσας διπλωματικής. Ο αναγνώστης μπορεί να ανατρέξει στο κεφάλαιο 3.3.1 για να ενημερωθεί για την πληρότητα των δεδομένων και πώς αυτή επηρέασε το πλήθος των ατυχημάτων που είχαμε στην διάθεση μας. Συνολικά υπήρχε 100% πληρότητα των δεδομένων σε 602 από τα ατυχήματα τα οποία και εισάχθηκαν στον αλγόριθμο σε αυτήν την πρώτη προσπάθεια εκπαίδευσης του.

#### Δεύτερη Δοκιμή

Η ακρίβεια της πρώτης δοκιμής υπολογίστηκε στα 51.5%. Η χαμηλή τιμή της ακρίβειας αυτής οφείλεται κατά κύριο λόγο στον μικρό αριθμό ατυχημάτων τον οποίο έχουμε στην διάθεση μας. Αυτός ο περιορισμένος αριθμός ατυχημάτων καθιστά αδύνατη την δημιουργία ενός αλγορίθμου με ικανοποιητική δυνατότητα πρόβλεψης όταν έχουμε 7 πιθανά responses και 6 observers με μία πληθώρα πιθανών τιμών το καθένα. Θα προσπαθήσουμε περιορίζοντας το πλήθος των responses και των πιθανών τιμών των observer να δημιουργήσουμε έναν αλγόριθμο μικρότερης εμβέλειας αλλά μεγαλύτερης ακρίβειας. Δηλαδή ο αλγόριθμος δεν θα έχει την δυνατότητα κάλυψης όλου του στόλου αλλά μόνο ενός κομματιού αυτού καθώς επίσης δεν θα μπορεί να προβλέψει όλα τα είδη ατυχημάτων αλλά μόνο μερικά. Στην πρώτη προσπάθεια να περιορίσουμε το εύρος λειτουργίας και να

αυξήσουμε την ακρίβεια θα αφαιρέσουμε τα 2 πιθανά responses “Foundered” & “War Hostilities”.

### Τρίτη Τέταρτη και Πέμπτη Δοκιμή

Η αφαίρεση των 2 πιθανών responses δεν προκάλεσε αισθητή μεταβολή στο βέλτιστο Accuracy αλλά μετέβαλε τον βέλτιστο αλγόριθμο. Σε αυτό το σημείο θα προσπαθήσουμε να βελτιώσουμε την δυνατότητα πρόβλεψης του αλγορίθμου περιορίζοντας τις πιθανές εκβάσεις των observers. Αναλυτικότερα θα περιορίσουμε τις εκβάσεις του Flag, Zone και Country Built διαδοχικά. Για τον περιορισμό των εκβάσεων αφαιρέθηκαν όλες οι εκβάσεις οι οποίες είχαν λιγότερες από 10 καταγραφές. Για παράδειγμα αν μία σημαία είχε αναμειχθεί σε λιγότερα από 10 ατυχήματα αυτά τα ατυχήματα αφαιρούνταν από το αρχείο εισαγωγής στο πρόγραμμα.

### Έκτη Δοκιμή

Στην Έκτη και τελευταία δοκιμή που πραγματοποιήθηκε αφαιρέθηκαν και 2 ακόμα responses “Contact” & “Fire/Explosion”. Αυτή ήταν και η τελευταία προσπάθεια που έγινε περιορισμού των δεδομένων για την δημιουργία ενός αλγορίθμου με ικανοποιητική δυνατότητα πρόβλεψης. Στον Πίνακα 16 φαίνονται τα αποτελέσματα των δοκιμών εκπαίδευσης του αλγορίθμου.

*Πίνακας 16 Αποτελέσματα δοκιμών εκπαίδευσης αλγορίθμου*

Αλγόριθμοι / Δοκιμες	1η	2η	3η	4η	5η	6η
Complex Tree	49,3%	49,0%	52,6%	56,9%	67,5%	<b>74,5%</b>
Medium Tree	51,2%	51,0%	52,4%	60,6%	<b>68,3%</b>	<b>74,5%</b>
Simple Tree	51,0%	52,4%	54,1%	<b>60,9%</b>	67,1%	71,9%
Linear SVM	34,7%	35,4%	36,5%	37,6%	43,8%	47,4%
Quadratic SVM	33,9%	34,7%	35,4%	36,2%	40,6%	45,3%
Cubic SVM	33,6%	34,0%	35,4%	35,6%	41,4%	45,3%
Fine Gaussian SVM	33,6%	34,0%	34,9%	36,2%	41,8%	43,8%
Medium Gaussian SVM	35,0%	35,7%	36,7%	38,8%	43,8%	45,8%
Coarse Gaussian SVM	35,0%	35,7%	37,6%	39,1%	43,4%	46,9%
Boosted Trees	50,8%	<b>52,9%</b>	<b>54,6%</b>	59,8%	67,1%	70,8%
Bagged Trees	<b>51,5%</b>	52,4%	<b>54,6%</b>	59,5%	<b>68,3%</b>	71,9%
RUSBoosted Trees	30,2%	42,9%	43,7%	46,4%	66,3%	66,7%
Observations	602	594	458	343	249	192
Response Classes	7	5	5	5	5	3

## Κεφάλαιο 5.4 Εφαρμογή και αξιολόγηση τελικού αλγορίθμου

Στο Κεφάλαιο 5.3.2 εκπαιδεύσαμε έναν αλγόριθμο με ικανότητα έγκυρης πρόβλεψης τύπου ατυχημάτων της τάξης του 74,5%. Για την αύξηση της ακρίβειας του αλγορίθμου το πλήθος των πιθανών εισαγωγών στο πρόγραμμα μειώθηκε αισθητά καθώς και το πλήθος των παραγόμενων αποτελεσμάτων. Συγκεκριμένα οι αποδεκτές εισαγωγές είναι:

Σημεία πλοίου: Ιαπωνία, Παναμάς, Σιγκαπούρη, Λιβερία, Ηνωμένες Πολιτείες Αμερικής

Ζώνη ατυχήματος: China Japan & Korea, South China & East Indies, Br.Isles N.Sea E.Chnl Biscay, Gulf Of Mexico

Ηλικία πλοίου: 0 έως 22 ηλικιακά έτη

Χώρα κατασκευής: Ιαπωνία, Νορβηγία, Ηνωμένες Πολιτείες Αμερικής

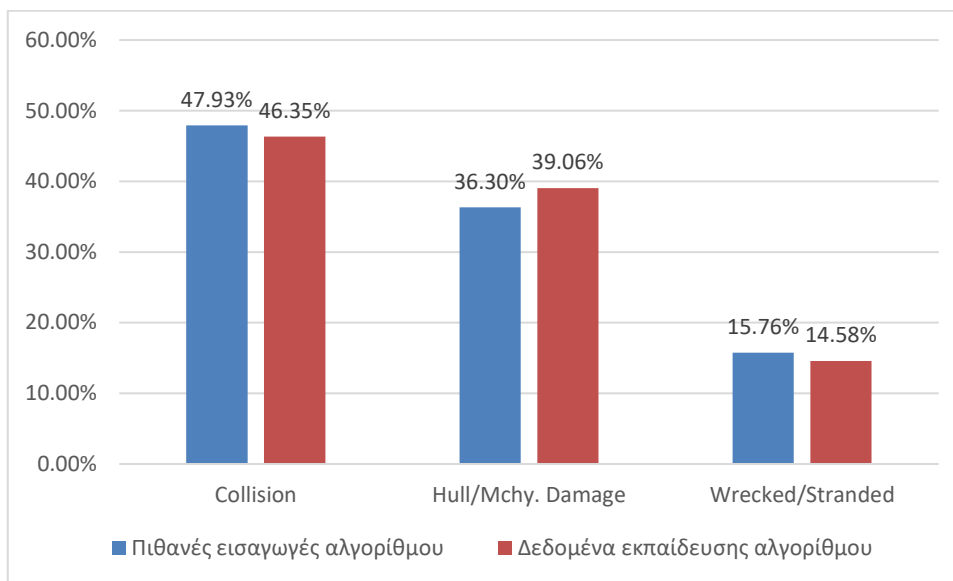
Είδος πλοίου: LNG Tanker, LPG Tanker

Παράλληλα ο αλγόριθμος μπορεί να επιστρέψει μόνο 3 από τα 7 πιθανά είδη ατυχημάτων: Collision, Hull&Mchy Damage και Wrecked/Stranded.

Στην προσπάθεια να γίνει μία εφαρμογή και παράλληλα αξιολόγηση του αλγορίθμου που αναπτύχθηκε ακολουθήθηκε η εξής διαδικασία. Παράχθηκε ένας πίνακας ο οποίος αποτελείται από όλους τους πιθανούς συνδυασμούς εισαγωγών που μπορούν να προκύψουν. Συγκεκριμένα υπάρχουν συνολικά 2760 πιθανοί συνδυασμοί στοιχείων που το πρόγραμμα μπορεί να επεξεργαστεί. Για να παραχθεί αυτός ο πίνακας δημιουργήθηκε ένα κομμάτι κώδικα VBA στο πρόγραμμα Microsoft Excel. Ο κώδικας που αναπτύχθηκε παρατίθεται στο Παράρτημα Α. Ο πίνακας αυτός στην συνέχεια χρησιμοποιήθηκε από τον εκπαιδευμένο αλγόριθμο για να προσδιοριστεί το είδος ατυχήματος σε όλες τις πιθανές εισαγωγές. Οι κατανομές του είδους ατυχήματος που προέκυψαν τελικά συγκρίθηκαν με τις κατανομές των ατυχημάτων που συγκρίθηκαν για την εκπαίδευση του αλγορίθμου. Έτσι μας δόθηκε η δυνατότητα να ελέγξουμε κατά πόσο ο αλγόριθμος είναι ακριβείς και η κατανομή των ατυχημάτων που προβλέπει ακολουθεί την κατανομή των ατυχημάτων που έχουν συμβεί.



## Στατιστική Ανάλυση Ναυτικών Ατυχημάτων σε πλοία LNG-LPG και Ανάπτυξη Αλγορίθμου Εξόρυξης Δεδομένων



Σχήμα 41 Σύγκριση κατανομής ατυχημάτων μεταξύ δεδομένων εκπαίδευσης και πιθανών εισαγωγών αλγορίθμου

Στο Σχήμα 41 παρουσιάζεται η κατανομή των ατυχημάτων ανά τύπο ατυχήματος για τα δεδομένα εκπαίδευσης του αλγορίθμου αλλά και για το εύρος των πιθανών εισαγωγών στον αλγόριθμο. Και στις τρεις κατηγορίες ατυχημάτων τα ποσοστά παρουσιάζουν πολύ μικρές αποκλίσεις μεταξύ των δεδομένων που χρησιμοποιήθηκαν για την εκπαίδευση του αλγορίθμου και των πιθανών εισαγωγών στον αλγόριθμο. Παρά το γεγονός ότι χρησιμοποιήθηκαν μόλις 192 ατυχήματα για την εκπαίδευση του αλγορίθμου ο ίδιος μπορεί να δώσει ακριβή αποτελέσματα για 2760 πιθανά σενάρια χωρίς να αποκλίνει αισθητά από την κατανομή των δεδομένων. Το αποτέλεσμα αυτό οφείλεται σε μεγάλο βαθμό στον περιορισμό των εκβάσεων και των predictors τον οποίο εφαρμόσαμε. Θα ήταν αδύνατο να πετύχουμε τέτοια ακρίβεια αν δεν είχαμε περιορίσει τις δυνατές τιμές των δεδομένων και τα δυνατά αποτελέσματα.

Η σελίδα αυτή είναι σκόπιμα κενή.

## ΚΕΦΑΛΑΙΟ 6. ΣΥΜΠΕΡΑΣΜΑΤΑ ΚΑΙ ΠΡΟΤΑΣΕΙΣ ΓΙΑ ΠΕΡΑΙΤΕΡΩ ΜΕΛΕΤΗ

Η πληρότητα των δεδομένων η οποία υπήρχε διαθέσιμη προς επεξεργασία σε πολλές περιπτώσεις ήταν αρκετά ελλιπής. Συγκεκριμένα πληροφορίες όπως η χώρα κατασκευής του πλοίου και το είδος του πλοίου που συμμετείχε στο ατύχημα παρουσίαζαν ελλείψεις που ξεπερνούσαν το 70%. Ταυτόχρονα πολλές πληροφορίες όπως ο καιρός την ώρα του ατυχήματος, η εθνικότητα του πληρώματος κ.α. απουσίαζαν εντελώς. Η έλλειψη αυτών των πληροφοριών οδηγεί σε δύο προβλήματα. Πρώτον, ένα μεγάλο κομμάτι του στατιστικού μας δείγματος έχει ελλιπής πληροφορίες και συνεπώς πρέπει να αφαιρεθεί από την μελέτη σε πολλές περιπτώσεις. Περιορίζοντας το μέγεθος του δείγματος που μελετάται περιορίζεται και η ακρίβεια των αποτελεσμάτων που λαμβάνονται. Δεύτερον, έχοντας διαθέσιμους μερικούς από τους παράγοντες οι οποίοι συμβάλουν στην ύπαρξη κάποιου ατυχήματος δεν μπορούμε να εξάγουμε με την μέγιστη δυνατή ακρίβεια κάποιο μοντέλο πρόβλεψης ατυχημάτων ή να βρούμε σχέσεις αιτίου-αποτελέσματος καθώς οι πληροφορίες που έχουμε για τα αίτια είναι ελλιπείς.

Παρά τις ελλείψεις που παρουσίαζαν τα δεδομένα μας καταφέραμε με ικανοποιητική ακρίβεια να προσδιορίσουμε 5 παράγοντες οι οποίοι επηρεάζουν τον τύπο ατυχήματος χρησιμοποιώντας τον στατιστικό έλεγχο  $\chi^2$ . Στην συνέχεια παρήχθησαν οι κατανομές ατυχημάτων συναρτήσει αυτών των πέντε παραγόντων για κάθε ένα τύπο ατυχήματος καθώς και οι κατανομή του στόλου συναρτήσει των διαφόρων παραγόντων και η συνολική κατανομή ατυχημάτων. Έχοντας τις κατανομές στόλου και τις συνολικές κατανομές ατυχημάτων σε αντιπαραβολή με τις κατανομές για κάθε είδος ατυχήματος και κάθε παράγοντα γνωρίζουμε πως ο κάθε παράγοντας επηρεάζει τον τύπο του ατυχήματος και έχουμε επισημάνει σε κάθε κατηγορία τις αποκλίσεις που εμφανίζονται.

Συγκεκριμένα η συχνότητα ύπαρξης ατυχημάτων **Collision** αυξάνεται σημαντικά όταν το πλοίο είναι τύπου LPG Tanker έχει σημαία Ιαπωνίας ή Παναμά, πλέει στις ζώνες China Japan & Korea ή South China & East Indies, έχει κατασκευαστεί στην Ιαπωνία και ηλικιακά είναι μεταξύ του 3<sup>ου</sup> και 9<sup>ου</sup> έτους ή του 13<sup>ου</sup> και 25<sup>ου</sup> έτους.

Η συχνότητα ύπαρξης ατυχημάτων **Wrecked/Stranded** αυξάνεται σημαντικά όταν το πλοίο είναι τύπου LPG Tanker έχει σημαία Ιαπωνίας, Λιβερίας, Σιγκαπούρης ή Νήσων Μάρσαλ, πλέει στις ζώνες West Indies ή China Japan & Korea, έχει κατασκευαστεί στην Ιαπωνία και ηλικιακά βρίσκεται στην πρώτη δεκαετία λειτουργίας.

Η συχνότητα ύπαρξης ατυχημάτων **Hull & Machinery Damage** αυξάνεται σημαντικά όταν το πλοίο έχει σημαία Isle of Man, Μπαχάμες, Λιβερία ή Σιγκαπούρης, πλέει στην ζώνη Br. Isles N. Sea E.Chnl Biscay, έχει κατασκευαστεί στην Γερμανία, στην Γαλλία, στην Ιταλία, στη Νορβηγία, στο Ηνωμένο Βασίλειο ή στις Ηνωμένες πολιτείες της Αμερικής και ηλικιακά βρίσκεται στο 16<sup>ο</sup>, 20<sup>ο</sup> ή 26<sup>ο</sup> ηλικιακό του έτος.

Η συχνότητα ύπαρξης ατυχημάτων **Fire & Explosion** αυξάνεται σημαντικά όταν το πλοίο είναι τύπου LPG Tanker έχει σημαία Λιβερίας ή Παναμά, πλέει στις ζώνες China Japan & Korea, South China & East Indies ή Gulf of Mexico, έχει κατασκευαστεί στην Νότια Κορέα ή στην Κίνα και ηλικιακά βρίσκεται στο 1<sup>ο</sup> έτος λειτουργίας του.

Στην συνέχεια παράχθηκαν οι συγκεντρωτικές συχνότητες ύπαρξης ατυχημάτων. Για την παραγωγή αυτών των συχνοτήτων χρησιμοποιήθηκαν τα καταγεγραμμένα ατυχήματα της βάσης δεδομένων Seaweb και το Fleet at Risk που μας δόθηκε από την εταιρεία Clarkson. Από το διάγραμμα συχνοτήτων που παράχθηκε προκύπτει ότι από το 1997 και μετά που έγινε η μελέτη οι συχνότητες των ατυχημάτων για υγραεριοφόρα πλοία βαίνουν φθίνουσες. Συγκεκριμένα ενώ ο στόλος ακολουθεί μία γνησίως αύξουσα καμπύλη και έχει υπερδιπλασιαστεί τα τελευταία 20 χρόνια τα ατυχήματα ανά έτος κυμαίνονται σταθερά μεταξύ 20 έως 60. Άρα συνολικά φαίνεται ότι οι μεταφορές με υγραεριοφόρα πλοία γίνονται ολοένα και πιο ασφαλέστερες με την πάροδο των ετών.

Τέλος μελετήσαμε και την επίδραση την ηλικίας του πλοίου στην συχνότητα ύπαρξης ατυχημάτων. Για να επιτευχθεί αυτό η τελευταία εικοσαετία χωρίστηκε σε τέσσερις πενταετίες. Για κάθε μία από τις πενταετίες υπολογίστηκε το Fleet at Risk και τα ατυχήματα που συνέβησαν για κάθε ηλικία. Από το διάγραμμα συχνοτήτων που παράχθηκε προέκυψαν τα εξής συμπεράσματα. Ποσοτικά με την πάροδο των ετών οι υψηλές συχνότητες της τάξης του 5E10-3 που εμφανίζονται στην πενταετία 1997-2001 σε μεγάλο ηλικιακό εύρος τείνουν να αντικατασταθούν στα ακόλουθα έτη από χαμηλές συχνότητες ατυχημάτων στο μεγαλύτερο κομμάτι της ζωής του πλοίου και να εμφανίζουν τοπικά μέγιστα σε συγκεκριμένες ηλικίες. Συγκεκριμένα για τις πενταετίες 2002-2006 και 2007-2011 το τοπικό μέγιστο των συχνοτήτων εμφανίζεται στα 25-29 ηλικιακά έτη ενώ για την τελευταία πενταετία 2012-2016 μετατοπίζεται στα 30-34 έτη. Άρα συγκεντρωτικά τα πλοία τα οποία ήταν συχνότερο να συμβεί ατύχημα στην δεκαετία 2002-2011 ήταν τα πλοία 25-29 ετών ενώ τα πλοία που είναι σήμερα συχνότερο να συμβεί ατύχημα είναι τα πλοία 30-34 ετών. Άρα με την πάροδο των ετών φαίνεται ότι τα υγραεριοφόρα πλοία στο σύνολο τους μπορούν να αξιοποιηθούν για μεγαλύτερο χρονικό διάστημα χωρίς να υπάρχει ιδιαίτερη ανησυχία για αυξημένη ύπαρξη ατυχημάτων. Αν και ιδιαίτερη προσοχή θα

πρέπει να δοθεί σε πλοία που φτάνουν το τριακοστό ηλικιακό έτος σήμερα καθώς η συχνότητες ατυχημάτων αυξάνονται δραματικά σε αυτό το σημείο.

Στο τελευταίο κομμάτι της μελέτης έγινε μία προσπάθεια να εκπαιδευτεί ένας αλγόριθμος με χρήση τεχνικών εξόρυξης δεδομένων. Συγκεκριμένα ο αλγόριθμος που εκπαιδεύτηκε έχει την δυνατότητα να προβλέψει τον τύπο ατυχήματος που είναι πιο πιθανό να συμβεί δοθέντων κάποιων παραμέτρων του πλοίου και δεδομένης της ύπαρξης κάποιου ατυχήματος. Με την χρήση αυτού του αλγορίθμου μας δίνεται η δυνατότητα να προβλέψουμε το είδος ατυχήματος που είναι πιο πιθανό να συμβεί σε ένα εύρος πλοίων χρησιμοποιώντας έναν περιορισμένο αριθμό ατυχημάτων που έχουν ήδη συμβεί. Συγκεκριμένα ο αλγόριθμος που εκπαιδεύτηκε χρησιμοποίησε 192 ατυχήματα και βρίσκεται σε θέση να προβλέψει επιτυχώς 2760 πιθανά σενάρια με ακρίβεια 74.5%. Λόγω των ελλείψεων που παρουσίαζαν τα δεδομένα που είχαμε στην διάθεση μας δεν ήταν δυνατόν να επεκτείνουμε περαιτέρω την ακρίβεια ή το εύρος εφαρμογής του αλγορίθμου.

Η μέθοδος εκπαίδευσης αλγορίθμων που περιεγράφηκε καθώς και ο αλγόριθμος που εκπαιδεύτηκε έχουν πολύ μεγάλα περιθώρια βελτίωσης και εξέλιξης. Συγκεκριμένα τα λίγα και ελλιπή δεδομένα που είχαμε στην διάθεση μας περιόρισαν σε μεγάλο βαθμό το εύρος λειτουργίας και την ακρίβεια του αλγορίθμου. Αν η ίδια μελέτη επαναληφθεί για άλλο τύπο πλοίου όπως τα Tanker η τα Bulk Carrier τα οποία διαθέτουν πολύ μεγαλύτερο αριθμό ατυχημάτων με χρήση κάποιας βάσης δεδομένων που παρέχει πλήρη και αναλυτικά δεδομένα για το κάθε ατύχημα όπως ο καιρός όταν έγινε το ατύχημα, η εθνικότητα του πληρώματος, η κατάσταση φόρτωσης του πλοίου, το σημείο που έγινε το ατύχημα ( λιμάνι, ανοιχτή θάλασσα κ.ο.κ. ) και άλλα είναι δυνατόν να δημιουργηθεί ένας αλγόριθμος με πολύ μεγάλη ακρίβεια πρόβλεψης και μεγάλο εύρος εφαρμογών.

Πέρα από αυτό θα μπορούσε να αναπτυχθεί και ένας αλγόριθμος ο οποίος να υπολογίζει την πιθανότητα ύπαρξης ή μη κάποιου ατυχήματος. Συγκεκριμένα αν κάποιος μπορούσε να συγκεντρώσει τα στοιχεία κίνησης των πλοίων παγκοσμίως ανεξάρτητα της ύπαρξης ατυχημάτων και τα ατυχήματα που έχουν συμβεί καθώς και όλους τους παράγοντες που τα επηρεάζουν θα είχε την δυνατότητα να εκπαιδεύσει έναν αλγόριθμο πρόβλεψης ατυχημάτων. Δηλαδή αντί να θεωρείται δεδομένο το ατύχημα και να προσπαθούμε να προβλέψουμε τον τύπο του ατυχήματος θα μας δινόταν η δυνατότητα να προβλέψουμε αν θα συμβεί ή όχι ένα ατύχημα. Περαιτέρω με χρήση αλγορίθμων Clusterer τους οποίους βέβαια δεν υποστηρίζει η εφαρμογή της Matlab R2016a που χρησιμοποιήθηκε στην παρούσα διπλωματική υπάρχει η δυνατότητα ανάπτυξης ενός αλγορίθμου ο οποίος θα έχει την δυνατότητα να υπολογίζει την πιθανότητα ύπαρξης κάθε τύπου ατυχήματος και την πιθανότητα μη ύπαρξης ατυχήματος για κάθε πλοίο για κάθε ταξίδι του. Η

δημιουργία ενός τέτοιου εργαλείου θα είχε αδιαμφισβήτητη αξία και χρήση όχι μόνο σε ερευνητικό επίπεδο αλλά και σε πρακτικό.

## ΒΙΒΛΙΟΓΡΑΦΙΑ

- [1] Witten, I. H., & Frank, E. (2000). *Data Mining Practical Machine Learning Tools and Techniques with Java Implementations*. San Francisco, CA: Morgan Kaufmann Publishers.
- [2] Han, J., & Kamber, M. (2001). *Data Mining Concepts and Techniques*. San Francisco, CA: Morgan Kaufmann Publishers.
- [3] Maimon, O., & Rokach, L. (2005). *The Data Mining and Knowledge Discovery Handbook*. New York, NY: Springer Science + Business Media Inc.
- [4] Fisher RA. The use of multiple measurements in taxonomic problems. *Ann Eugen* 1936, 7:179–188
- [5] Fielding A, O’Muirheartaigh CA. Binary segmentation in survey analysis with particular reference to AID. *The Statistician* 1977, 25:17–28.
- [6] Messenger R, Mandell L. A modal search technique for predictive nominal scale multivariate analysis. *J Am Stat Assoc* 1972, 67:768–772.
- [7] Morgan JN, Sonquist JA. Problems in the analysis of survey data, and a proposal. *J Am Stat Assoc* 1963, 58:415–434.
- [8] Quinlan JR. Learning with continuous classes. *Proceedings of the 5th Australian Joint Conference on Artificial Intelligence* 1992, 343–348.
- [9] Loh WY. Regression trees with unbiased variable selection and interaction detection. *Stat Sin* 2002, 12:361–386.
- [10] N. Cristianini and J. Shawe-Taylor. *An Introduction to Support Vector Machines and Other Kernel-based Learning Methods*, Cambridge University Press, 2000. © 2009 by Taylor & Francis Group, LLC  
*References* 57
- [11] S. Haykin. *Neural Networks: A Comprehensive Foundation*, Tsinghua University Press, 2001.
- [12] R. Polikar, Ensemble based systems in decision making, *IEEE Circuits and Systems Magazine*, 6 (3) (2006), pp. 21–45
- [13] E.K. Tang, P.N. Suganthan, X. Yao, An analysis of diversity measures, *Machine Learning*, 65 (2006), pp. 247–271
- [14] L.I. Kuncheva, C.J. Whitaker, Measures of diversity in classifier ensembles and their relationship with the ensemble accuracy, *Machine Learning*, 51 (2) (2003), pp. 181–207

- [15] A Basic Introduction to Neural Networks.  
<<http://www.pages.cs.wisc.edu/~bolo/shipyard/neural/local.html> > (accessed August 2012).
- [16] MathWorks, Support Vector Machines (SVM).  
<<http://www.mathworks.com.au/help/toolbox/bioinfo/ug/bs3tbev -1.html>>  
(accessed August 2012).
- [17] A. Rahman and B. Verma, Ensemble Classifier Generation using Non-uniform Layered Clustering and Genetic Algorithm,  
Elsevier Knowledge Based Systems, vol. 43 (May 2013), pp. 30– 42, 2013.
- [18] R. Maclin, J.W. Shavlik, Combining the predictions of multiple classifiers: using competitive learning to initialize neural networks, in: International Joint Conference on Artificial Intelligence, 1995, pp. 524–531.
- [19] T. Yamaguchi, K.J. Mackin, E. Nunohiro, J.G. Park, K. Hara, K. Matsushita, M. Ohshiro, K. Yamasaki, Artificial neural network ensemble-based land-cover classifiers using MODIS data, *Artificial Life and Robotics*, 13 (2) (2009), pp. 570–574.
- [20] H. Chen, X. Yao, Regularized negative correlation learning for neural network ensembles, *IEEE Transactions on Neural Networks*, 20 (12) (2009), pp. 1962–1979.
- [21] H. Chen, X. Yao, Multiobjective neural network ensembles based on regularized negative correlation learning, *IEEE Transactions on Knowledge and Data Engineering*, 22 (12) (2010), pp. 1738–1751.
- [22] T.K. Ho, The random subspace method for constructing decision forests, *IEEE Transaction on Pattern Analysis and Machine Intelligence*, 20 (8) (1998), pp. 832–844.
- [23] A. Bertoni, R. Folgieri, G. Valentini, Bio-molecular cancer prediction with random subspace ensembles of support vector machines, *Neurocomputing*, 63 (2005), pp. 535– 539.
- [24] L.I. Kuncheva, J.J. Rodriguez, C.O. Plumpton, D.E. Linden, S.J. Johnston, Random subspace ensembles for FMRI classification, *IEEE Transaction on Medical Imaging*, 29 (2) (2010), pp. 531–542.
- [25] G. Martínez-Muñoz, A. Sánchez-Martínez, D. HernándezLobato, A. Suarez, Class-switching neural network ensembles, *Neurocomputing*, 7 (2008), pp. 2521–2528.
- [26] T.G. Dietterich, G. Bakiri, Solving multiclass learning problems via error-correcting output codes, *Journal of Artificial Intelligence Research*, 2 (1995), pp. 263–286
- [27] L. Rokach, O. Maimon, I. Lavi, Space decomposition in data mining: a clustering approach, in: *International Symposium on Methodologies for Intelligent Systems*, 2003, pp. 24–31.
- [28] J. Xiuping, J.A. Richards, Cluster-space classification: a fast k-nearest neighbour classification for remote sensing



- hyperspectral data, in: IEEE Workshop on Advances in Techniques for Analysis of Remotely Sensed Data, 2003, pp. 407–410.
- [29] L.I. Kuncheva, Cluster-and-selection method for classifier combination, in: International Conference on Knowledge-Based Intelligent Engineering Systems and Allied Technologies (KES), 2000, pp. 185–188.
- [30] B. Tang, M.I. Heywood, M. Shepherd, Input partitioning to mixture of experts, in: International Joint Conference on Neural Networks, 2002, pp. 227–232.
- [31] G. Nasierding, G. Tsoumakas, A.Z. Kouzani, Clustering based multi-label classification for image annotation and retrieval, in: IEEE International Conference on Systems, Man and Cybernetics, 2009, pp. 4514–4519.
- [32] S. Eschrich, L.O. Hall, Soft partitions lead to better learned ensemble, in: Annual meeting of the North American fuzzy information processing society (NAFIPS), 2002, pp. 406–411.
- [33] A. Rahman and B. Verma, A Novel Layered Clustering based Approach for Generating Ensemble of Classifiers, IEEE Transaction on Neural Networks, vol 22, no 5, pp 781– 792, 2011.
- [34] A. Rahman and B. Verma, A Novel Ensemble Classifier Approach using Weak Classifier Learning on Overlapping Clusters, Proc. IEEE International Joint Conference on Neural Networks (IJCNN), Barcelona, Spain, 2010.
- [35] A. Rahman and B. Verma, Influence of Unstable Patterns in Layered Cluster Oriented Ensemble Classifier, Proc. IEEE International Joint Conference on Neural Networks (IJCNN), Brisbane, Australia, 2012.
- [36] A. Rahman and B. Verma, Cluster Based Ensemble Classifier Generation by Joint Optimization of Accuracy and Diversity, International Journal of Computational Intelligence and Applications, vol. 12, no. 4, DOI: 10.1142/S1469026813400038, 2013.
- [37] A. Rahman and B. Verma, Cluster Oriented Ensemble Classifiers using Multi-Objective Evolutionary Algorithm, Proc. IEEE International Joint Conference on Neural Networks (IJCNN), pp. 829–834, Dallas, Texas, 2013.
- [38] A. Rahman, B. Verma, and X. Yao, Non-uniform Layered Clustering for Ensemble Classifier Generation and Optimality, 19th International Conference on Neural Information Processing (ICONIP 2012): Lecture Notes in Computer Science, vol. 6443, pp. 551–558, 2010.
- [39] L. Breiman, Bagging predictors, Machine Learning, 24 (2) (1996), pp. 123–140

- [40] L. Breiman, Random forests, *Machine Learning*, 45 (1) (2001), pp. 5–32
- [41] L. Breiman, Pasting small votes for classification in large databases and on-line, *Machine Learning*, 36 (1999), pp. 85–103
- [42] R.E. Schapire, The strength of weak learnability, *Machine Learning*, 5 (2) (1990), pp. 197–227
- [43] Y. Freund, R.E. Schapire, Decision-theoretic generalization of on-line learning and an application to boosting, *Journal of Computer and System Sciences*, 55 (1) (1997), pp. 119–139
- [44] MAN (2009) *Propulsion Trends in LNG Carriers Two-stroke Engines*. Available at: [http://www.mandiesel.com/files/news/files\\_of8074/5510-0035-01ppr.pdf](http://www.mandiesel.com/files/news/files_of8074/5510-0035-01ppr.pdf) (Accessed: 21 Feb 2010)
- [45] Ship Knowledge, GL 2005
- [46] Younglove, B. A., and James F. Ely. "Thermophysical properties of fluids. II. Methane, ethane, propane, isobutane, and normal butane." *Journal of Physical and Chemical Reference Data* 16.4 (1987): 577-798.
- [47] ABS Consulting Inc., 2004, Consequence Assessment Methods for Incidents Involving Releases from Liquefied Natural Gas Carriers, Report 131-04,
- [48] Bagdonavicius, V., Kruopis, J., Nikulin, M.S. (2011). "Non-parametric tests for complete data", ISTE & WILEY: London & Hoboken.
- [49] Geisser, S.; Johnson, W.M. (2006) *Modes of Parametric Statistical Inference*, John Wiley & Sons, [ISBN 978-0-471-66726-1](https://doi.org/10.1002/9780471667261)
- [50] B. S. Everitt: *The Cambridge Dictionary of Statistics*, [Cambridge University Press](https://doi.org/10.1017/C9780521690277), Cambridge (3rd edition, 2006). [ISBN 0-521-69027-7](https://doi.org/10.1017/C9780521690277)
- [51] [Greenwood, P.E.](https://doi.org/10.1002/978047155779X), Nikulin, M.S. (1996) *A guide to chi-squared testing*. Wiley, New York. [ISBN 0-471-55779-X](https://doi.org/10.1002/978047155779X)
- [52] Nikulin, M.S. (1973). "Chi-squared test for normality". In: *Proceedings of the International Vilnius Conference on Probability Theory and Mathematical Statistics*, v.2, pp. 119–122.
- [53] Bagdonavicius, V., Nikulin, M.S. (2011) "Chi-squared goodness-of-fit test for right censored data". *The International Journal of Applied Mathematics and Statistics*, p. 30-50.
- [54] Lloyd's Register (<http://www.lr.org/en/>)
- [55] Watanabe et Al, 2007

## ΠΑΡΑΡΤΗΜΑ Α – ΚΩΔΙΚΑΣ VBA

```
Sub Button1_Click()  
  
Dim ii As Integer  
  
Dim jj As Integer  
  
Dim ll As Integer  
  
Dim kk As Integer  
  
Dim XX As Integer  
  
Dim k As Integer  
  
k = 1  
  
For ii = 2 To 6  
  
    For jj = 2 To 5  
  
        For ll = 2 To 24  
  
            For kk = 2 To 4  
  
                For XX = 2 To 3  
  
                    Cells(k, 1).Value = Worksheets(1).Cells(ii, 1).Value  
  
                    Cells(k, 2).Value = Worksheets(1).Cells(jj, 2).Value  
  
                    Cells(k, 3).Value = Worksheets(1).Cells(ll, 3).Value  
  
                    Cells(k, 4).Value = Worksheets(1).Cells(kk, 4).Value  
  
                    Cells(k, 5).Value = Worksheets(1).Cells(XX, 5).Value  
  
                    k = k + 1  
  
                Next XX  
  
            Next kk  
  
        Next ll  
  
    Next jj  
  
Next ii  
  
End Sub
```