

Ελλιπείς τιμές (missing values)

1	Gender	Race	Birth_Ye...	Marital_...	Years_o...	Hours_P...	Preferre...	Preferre...	Preferre...	Read_N...	Online_...	Online_...	Facel
2	M	White	1972	M	8	1	Firefox	Google	Yahoo	Y	N	N	Y
3	M	Hispanic	1981	S	14	2	Chrome	Google	Hotmail	Y	N	N	Y
4	F	African A...	1977	S	6	2	Firefox	Yahoo	Yahoo	Y	Y		Y
5	F	White	1961	D	8	6	Firefox	Google	Hotmail	N	Y	N	N
6	M	White	1954	M	2	3	Internet ...	Bing	Hotmail	Y	Y	N	Y
7	M	African A...	1982	D	15	4	Internet ...	Google	Yahoo	Y	N	Y	N
8	M	African A...	1981	D	11	2	Firefox	Google	Yahoo		Y		Y
9	M	White	1977	S	3	3	Internet ...	Yahoo	Yahoo	Y			Y
10	F	African A...	1969	M	6	2	Firefox	Google	Gmail	N	Y	N	N
11	M	White	1987	S	12	1	Safari	Yahoo	Yahoo	Y		Y	Y
12	F	Hispanic	1959	D	12	5	Chrome	Google	Gmail	Y	N	N	Y

Ελλιπείς
τιμές

Ελλιπείς τιμές (missing values)

Προκειμένου να διορθώσουμε τις εγγραφές που παρουσιάζουν ελλιπείς τιμές μπορούμε να εφαρμόσουμε μια από τις παρακάτω μεθόδους

- ◆ Αντικατάσταση τιμής βάσει
 - ◆ του ΜΟ των υπολοίπων τιμών
 - ◆ Της μικρότερης τιμής
 - ◆ Της μεγαλύτερης τιμής
 - ◆ Να βάλουμε μια δική μας τιμή
 - ◆ Να βάλουμε το μηδέν.

Δεδομένα με ακραίες τιμές (outliers)

Μέσω της διαδικασίας «Outlier detection» μπορούμε να εντοπίσουμε αντικείμενα τα οποία παρουσιάζουν συμπεριφορά διαφορετική από την αναμενόμενη.

Σε αυτά τα αντικείμενα συνήθως εστιάζει ένας αναλυτής για περαιτέρω ανάλυση γιατί ενδεχομένως

- ◊ Να δείχνουν μια οικονομική απάτη
- ◊ Να ανιχνεύουν κάποιο φαινόμενο σε μια ανάλυση ιατρικών δεδομένων
- ◊ Να ανιχνεύουν παραβιάσεις στην ασφάλεια ενός δικτύου.

Βέβαια υπάρχουν και περιπτώσεις που μπορεί απλά να προστέθηκαν από λάθος στο δείγμα και να πρέπει να απομακρυνθούν για διορθωθεί τα δεδομένα.

Μεγάλα Δεδομένα στη Ναυτιλία

- Η ναυτιλιακή βιομηχανία είναι ένα σύνθετο σύστημα που απαιτεί

γρήγορη προσαρμογή στις αλλαγές συνθηκών για στις οποίες η λήψη αποφάσεων πρέπει να λαμβάνει υπόψη

μεγάλο αριθμό παραμέτρων.



Μεγάλα Δεδομένα στη Ναυτιλία

- Το θαλάσσιο εμπόριο αντιπροσωπεύει πάνω από το 90 τοις εκατό του παγκόσμιου εμπορίου ως προς τον όγκο. Λόγω του μεγέθους του δικτύου που δημιουργείται, οι εταιρείες ναυτιλιακής εφοδιαστικής αλυσίδας αντιμετωπίζουν προβλήματα μεγάλης κλίμακας σχεδιασμού σε στρατηγικό, τακτικό και επιχειρησιακό επίπεδο.
- Η διαδικασία λήψης αποφάσεων σχετικά με τη ναυτιλιακή οικονομική πρέπει να διασφαλίζει την ασφάλεια, να ελαχιστοποιεί το κόστος και να βελτιώνει τα μέσα παραγωγικότητας λαμβάνοντας υπόψη έναν **μεγάλο αριθμό ευαίσθητων παραμέτρων με υψηλή μεταβλητότητα**.
- Παρόλο που τα μεγάλα δεδομένα έχουν υποστηρίξει σημαντικά την αγορά και διάφορους βιομηχανικούς κλάδους, η υιοθέτησή τους από τη ναυτιλιακή βιομηχανία είναι ακόμη αργή

Μεγάλα Δεδομένα στη Ναυτιλία

- Η ανάπτυξη σύγχρονων συστημάτων πλοήγησης, αισθητήρων και συστήματων παρακολούθησης με τις πρόσφατες εξελίξεις στην τεχνολογία, ανοίγει για την η ναυτιλιακή βιομηχανία τον δρόμο για την προσέγγιση στα οφέλη της ψηφιακής εποχής.
- Ο αυξανόμενος όγκος των διαθέσιμων δεδομένων σχετικά με τις επιδόσεις και τη ναυσιπλοΐα του πλοίου φέρνει μία ευρεία γκάμα δυνατοτήτων, στη παρακολούθηση σκαφών σε πραγματικό χρόνο έως την εξαγωγή γνώσης μέσω ανάλυσης δεδομένων.
- Ο όγκος και η ποικιλία των θαλάσσιων δεδομένων καθιστά το ζήτημα ένα πρόβλημα Big Data

Μεγάλα Δεδομένα στη Ναυτιλία

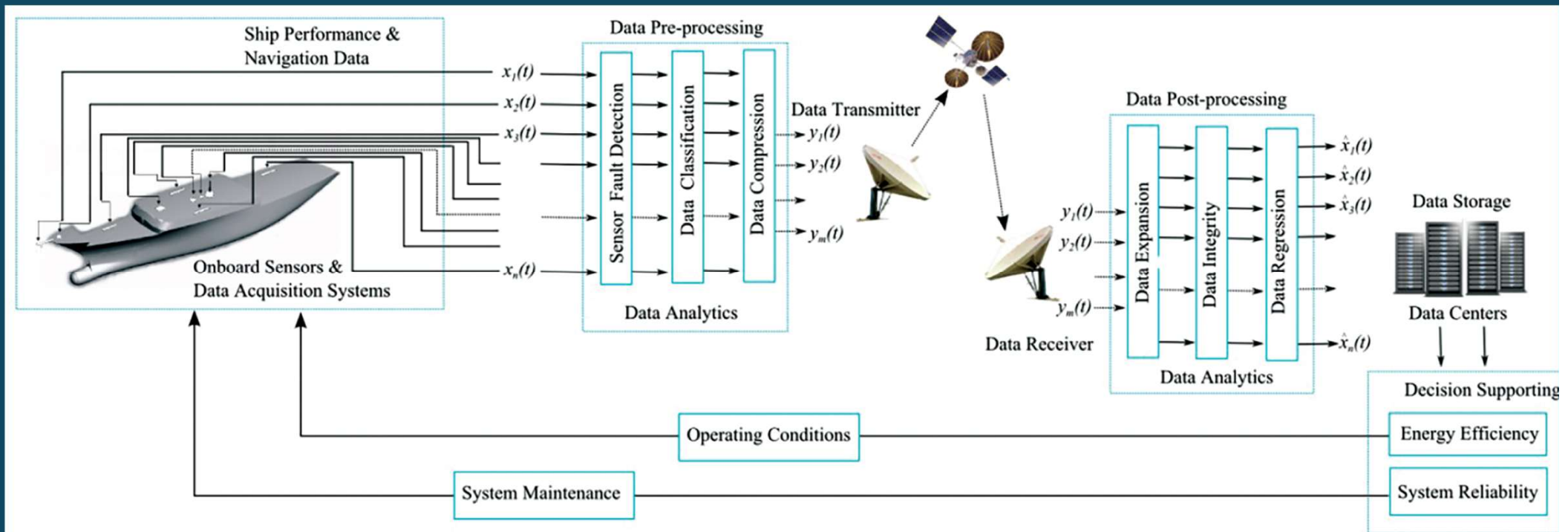
- Ένα υποσύνολο δεδομένων επιδόσεων και πλοήγησης πλοίου, όπως ο μοναδικός αριθμός αναγνώρισης του σκάφους, η θέση, πορεία, ταχύτητα και προορισμός, μεταφέρονται από το Αυτόματο Σύστημα Αναγνώρισης (AIS).
- Πλοία σε διεθνή ταξίδια και όλα τα επιβατηγά πλοία υποχρεωτικά διαθέτουν συστήματα AIS μέσω των οποίων ανταλλάσσονται δεδομένα με άλλα κοντινά πλοία, επίγεια συστήματα και δορυφόρους υποστηρίζοντας σημαντικά και την αποφυγή ατυχημάτων.

Μεγάλα Δεδομένα στη Ναυτιλία

- Επιπλέον, το Voyage Data Recorder (VDR), το οποίο είναι απαραίτητο σε όλα τα επιβατηγά πλοία και άλλα πλοία με tonnage 3000 ή περισσότερο, συνδέεται με έναν αριθμό ηλεκτρονικών συσκευών και αποθηκεύει τις καταγεγραμμένες πληροφορίες κάθε ταξιδιού ως ηλεκτρονικά δεδομένα.
- Τα στοιχεία δεδομένων περιλαμβάνουν *τη θέση του πλοίου, την ταχύτητα, ήχο από μικρόφωνα γέφυρας, δεδομένα ραντάρ, βάθος νερού, ταχύτητα και κατεύθυνση ανέμου, δεδομένα από τους συναγερμούς, κατάσταση ανοιγμάτων γάστρας, επιτάχυνση και τάσεις κύτους, τάξη και απόκριση του κινητήρα και του πηδαλίου* κ.α.
- Ο όγκος των δεδομένων είναι τόσο μεγάλος που πολλές φορές κατά τη διάρκεια μεγάλων ταξιδιών, τα παλαιότερα δεδομένα πρέπει να αντικατασταθούν για αποθήκευση από νέα.

Μεγάλα Δεδομένα στη Ναυτιλία

- Αφού ολοκληρωθεί το ταξίδι, τα δεδομένα συνήθως απορρίπτονται. Ο κύριος σκοπός του VDR είναι η ανάλυση δεδομένων σε περίπτωση ατύχηματος.
- Ωστόσο, αντί να πετάμε αυτές τις πληροφορίες, θα μπορούσαν να επεξεργαστούν και να χρησιμοποιηθεί αποτελεσματικά με τη βοήθεια του τεχνικών ανάλυσης μεγάλων δεδομένων



ΤΥΠΟΙ ΠΗΓΩΝ ΚΑΙ ΔΕΔΟΜΕΝΩΝ ΣΤΗ ΝΑΥΤΙΛΙΑ

Δεδομένα πλοήγησης

Δεδομένα απόδοσης πλοίου

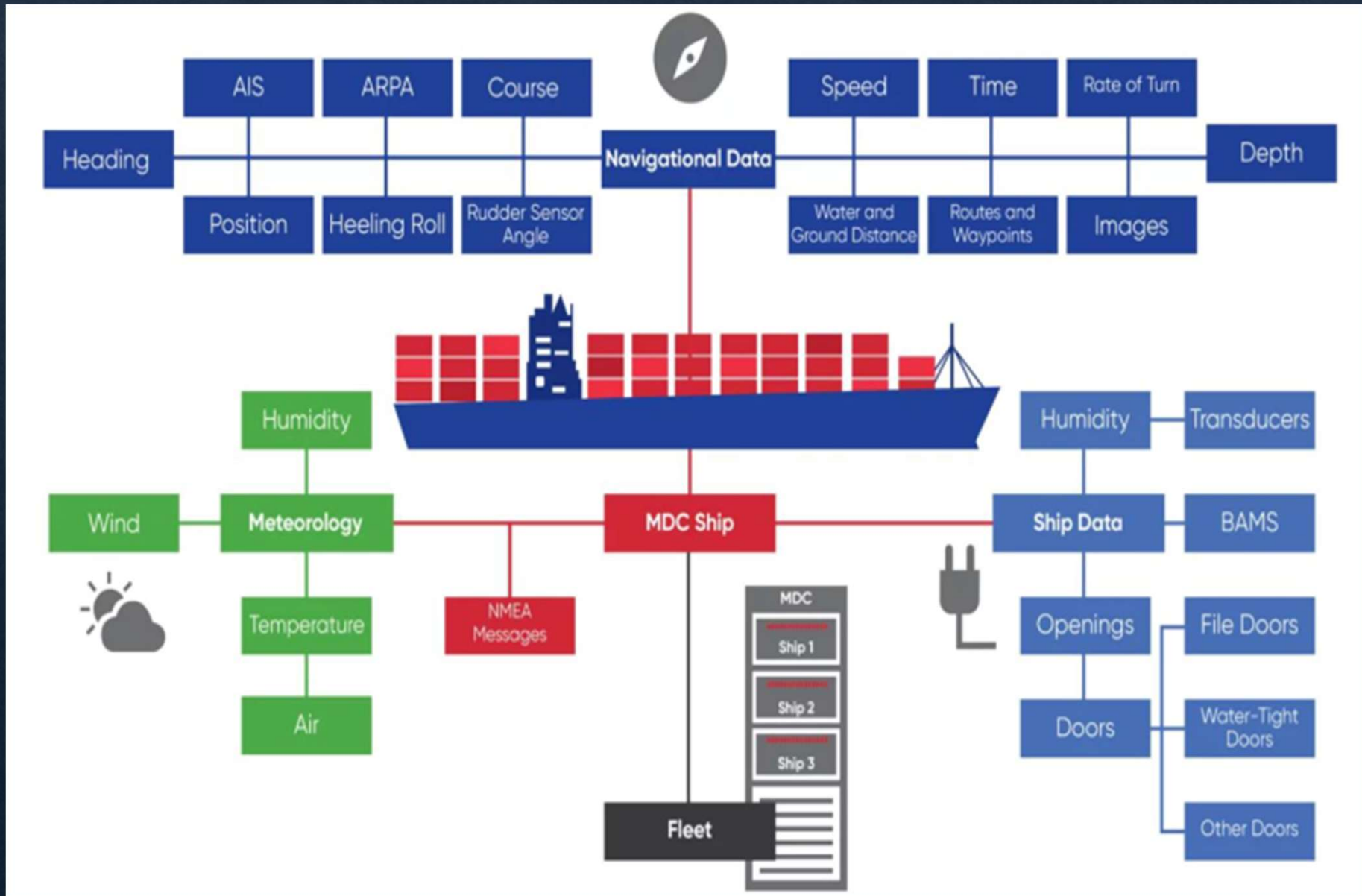
Δεδομένα καιρού

Δεδομένα φορτίου

Δεδομένα λιμένας

Ρυθμιστικά δεδομένα

Παραγωγή δεδομένων από το πλοίο



Μεγάλα Δεδομένα στη Ναυτιλία – Πηγές Δεδομένων

Port Operations Systems

- arrival/departure timestamps,
- cargo information,
- crew lists,
- customs declarations

The data tend to be structured or semi-structured in nature and fairly accurate as it is often used to calculate port fees and other commercial operations

AIS Stations

AIS is an automatic tracking system installed on vessels and used primarily by vessel traffic services (VTS).

AIS allows maritime stakeholders to monitor and track vessel movements in real time within the range of very high frequency (VHF) radios on base stations located along coastlines or satellites.

AIS data is transmitted by ships every 3–10 s encoded in a specialised format and contains information about the vessel:

- unique identification, name, flag, type, dimension
- the current status (longitude, latitude, speed, heading),
- voyage (e.g., destination, estimated time of arrival, draught).

It is estimated that over 520 million vessel positions are recorded and processed daily

Μεγάλα Δεδομένα στη Ναυτιλία – Πηγές Δεδομένων

Sensing Devices

Modern vessels use built-in sensors, advanced dynamic position systems, control systems, navigational sensor systems, etc., to enable automatic fault detection and preemptive maintenance, cargo tracking, and energy-efficient operations.

IoT-based sensors are also used in containers for monitoring and maintaining temperature and humidity to ensure the viability of perishable goods.

Oceanic and Weather Stations

Various decisions in the shipping sector, such as route planning and voyage optimisation, are directly impacted by oceanic and weather data

Such data typically includes temperature, humidity, wind speed, rainfall, sea currents, tidal variations, and wave characteristics.

Usually, the maritime industry forecasts weather information using ensemble models, historical data, and current measurements.

Μεγάλα Δεδομένα στη Ναυτιλία - Energy efficiency improvement

- Με στόχο την βελτίωση της ενεργειακής απόδοσης, μπορούν να χρησιμοποιηθούν πληροφορίες για
 - την απόδοση του πλοίου
 - την πλοήγηση και την ανάπτυξη στρατηγικών πλοήγησης
 - την παρακολούθηση κατανάλωσης καυσίμου,
 - διαφόρων εκπομπών ρύπων,
 - χρήσης φωτισμού και θέρμανσης

και παρόμοιες διαδικασίες μπορούν να οδηγήσουν σε ιδέες που υποστηρίζουν τη λήψη αποφάσεων.

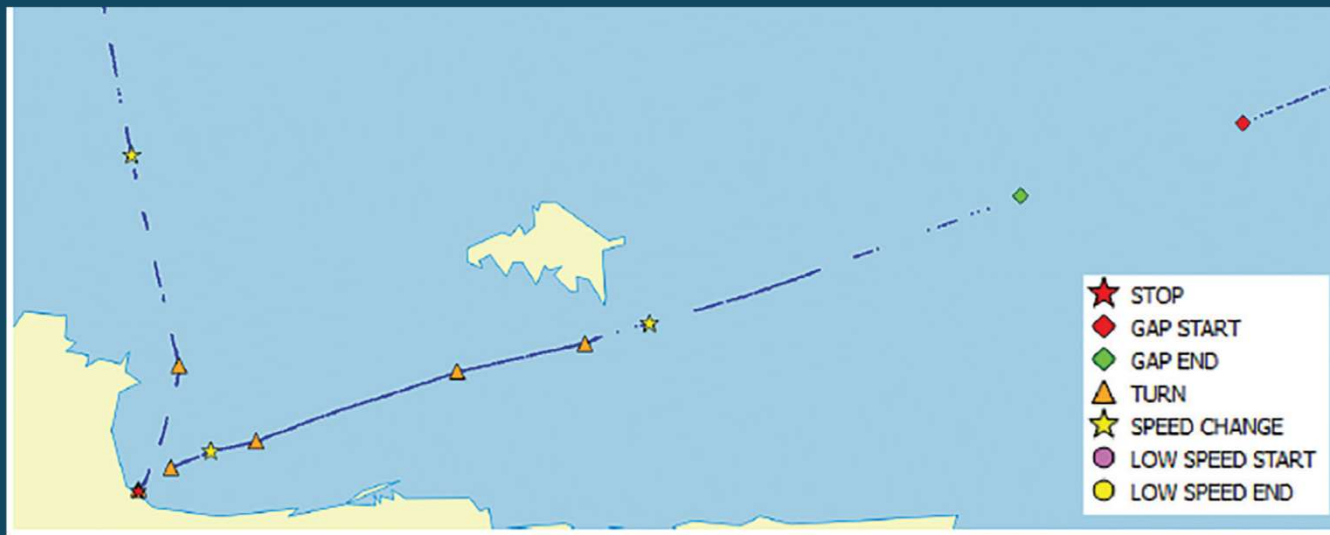
Μεγάλα Δεδομένα στη Ναυτιλία - Energy efficiency improvement

- Δεδομένα όπως
 - ταχύτητα και κατεύθυνση ανέμου,
 - μέσο βύθισμα,
 - ισχύς κύριου κινητήρα,
 - ταχύτητα άξονα και κατανάλωση καυσίμου

αναλύονται και αρκετές τάσεις υψηλότερης κατανάλωσης καυσίμου που επηρεάζονται κάτω από αυτές τις παραμέτρους προλαμβάνονται

Μεγάλα Δεδομένα στη Ναυτιλία – Safety

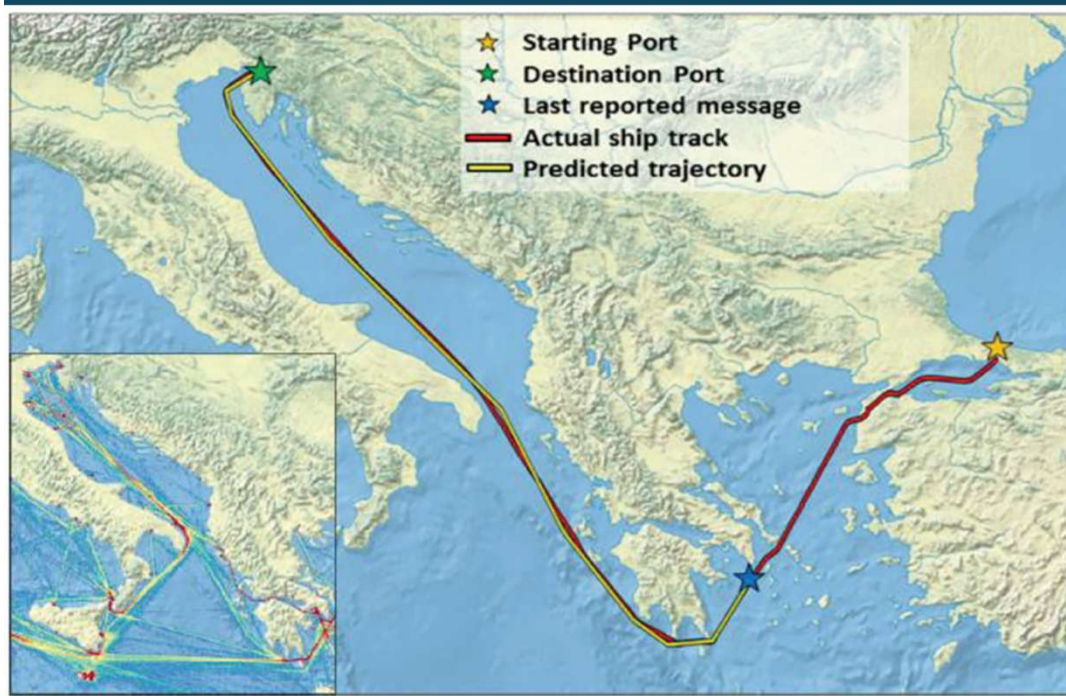
- Η ασφάλεια στη ναυσιπλοΐα μπορεί επίσης να επωφεληθεί από την εφαρμογή μεθόδων anomaly detection σε θαλάσσιες επιχειρήσεις
- Δεδομένα αισθητήρων, τα οποία περιλαμβάνουν πληροφορίες για το περιβάλλον και το σύστημα των πλοίων, παρέχονται και αναλύονται μέσω ενός Big Data συστήματος που χρησιμοποιεί Auto Associative Kernel Regression και τεχνικές δοκιμής διαδοχικής αναλογίας πιθανοτήτων για την ανίχνευση ανωμαλιών και κρίσιμων σημείων στην διαδρομή ενός πλοίου



Critical points identified
along a vessel trajectory

Μεγάλα Δεδομένα στη Ναυτιλία – Logistics optimization

- Όταν πρόκειται για τη μείωση των logistics και του λειτουργικού κόστους, τεχνικές Μεγάλων Δεδομένων μπορούν να χρησιμοποιηθούν για να ελαχιστοποιηθεί ο χρόνος των πλοίων που περνούν αγκυροβολημένοι έξω από το λιμάνι περιμένοντας μια διαθέσιμη υποδοχή, και να μεγιστοποιηθεί αντίστοιχα η χρήση των αποβάθρων
- Επίσης στην πρόγνωση τροχιάς πλοίου με βάση ιστορικά δεδομένα



vessel position prediction 48 hours
in advance based on historical
maritime traffic data

Accurate prediction of routes can
be used to better estimate times of
arrival in ports

Μεγάλα Δεδομένα στη Ναυτιλία – Application in Ports and Terminals

PORT NAME / PORT AUTHORITY	DESCRIPTION	PURPOSE / BENEFIT
Ports in Singapore and Malaysia	In Singapore and Malaysia, ports utilize big data techniques to create advanced inspection systems to assess the history and cargo type of importers.	The purpose is to segregate importers that most require inspection and allow other importers to operate quickly, without impacting the port's security objectives.
Maritime and Port Authority, Singapore	In August 2015, Singapore's Maritime and Port Authority (MPA) signed a Memorandum of Understanding (MOU) with IBM to establish a two-year agreement for big data. The agreement involved developing a platform using IBM's Traffic Prediction tool to forecast vessel arrival times and estimate potential traffic congestion using fusion analytics. The platform also relies on data mining and anomaly detection by using IBM Incident Detection Module and IBM System G. According to Goh Kwong Heng, CIO of MPA, the port authority plans to invest in big data to improve port operations and activities. The MPA aims to use data analytics platforms to complement its port management systems in detecting anomalies and supporting both operations and planning processes. The authority also plans to invest in technologies such as drones and mobile apps to manage marine accidents and improve the efficiency of its port workers.	The port uses the platform to improve productivity and marine safety at the major trans-shipment hub. For example, by collecting information on ship positions and weather data, the platform helps avoid accidents by inferring the most likely path ships would take in a given situation. This also helps prevent illegal bunkering by detecting unusual movement patterns. According to Andrew Tan, the CEO of MPA, the port authority's collaboration with IBM enabled a mix of research expertise, software technologies and maritime domain knowledge to create new capabilities for the maritime industry in Singapore.

Μεγάλα Δεδομένα στη Ναυτιλία – Application in Ports and Terminals

PORT NAME / PORT AUTHORITY	DESCRIPTION	PURPOSE / BENEFIT
Hamburg Port, Germany	The Hamburg Port in Germany uses a cloud-based analytics tool called SmartPort Logistics. The tool pulls different types of data, such as vessel positions, height and width of bridges, and planned driving routes. This data can be viewed in real-time on mobile applications.	The tool aims to streamline the flow of goods. It allows port workers to know precisely when ships are expected to dock, and truck drivers to know when cargo is expected to be offloaded.
Port of Cartagena, Columbia	The port uses Cisco and IBM solutions for IoT analytics. The solutions help forecast equipment failures.	The solutions ensure proper and timely maintenance of port machinery.
Port of Rotterdam, The Netherlands	The Port of Rotterdam has recently implemented big data analytics.	Sustainable supply chain analysis for Reefer containers.

Μεγάλα Δεδομένα στη Ναυτιλία – Περιορισμοί και Προκλήσεις

- Όπως συμβαίνει σε όλα τα συστήματα που περιλαμβάνουν Big Data εμφανίζονται προκλήσεις που σχετίζονται με την ποιότητα και την ποσότητα των δεδομένων.
- Η απόδοση και πλοήγηση πλοίου σχετίζεται με δεδομένα που συλλέγονται από αισθητήρες και συστήματα απόκτησης δεδομένων τα οποία μπορεί να έχουν αστοχίες που οφείλονται σε σφάλματα αισθητήρα ή τυχαία λάθη που έγιναν κατά τη διάρκεια της καταγραφής.
- Μερικές φορές, τα σφάλματα μπορεί να είναι εσκεμμένα, όταν οι χειριστές έχουν οικονομικά συμφέροντα στην αναφορά παραπλανητικών στοιχείων ή σε μια προσπάθεια συγκάλυψης παράνομης δραστηριότητας.
- Μια πιθανή λύση είναι η αυτοματοποιημένη εισαγωγή δεδομένων, η οποία όμως μπορεί να είναι πολύ δαπανηρή

Μεγάλα Δεδομένα στη Ναυτιλία – Περιορισμοί και Προκλήσεις

- Για να ελαχιστοποιηθούν τα σφάλματα, τα δεδομένα πρέπει να περάσουν από ποιοτικό έλεγχο πριν αναλυθούν.
- Ο έλεγχος - καθαρισμός δεδομένων περιλαμβάνει εργασίες όπως η ανίχνευση ανωμαλιών, ανίχνευση αντικρουόμενων τιμών και η αντιμετώπιση της ελλιπούς συμπλήρωσης δεδομένων.
- Ωστόσο, στην βιβλιογραφία επισημαίνεται ότι σε πολλές περιπτώσεις οι στρατηγικές καθαρισμού δεν είναι κατάλληλες για Big Data και ο θόρυβος και οι ακραίες τιμές που απορρίπτονται σε αυτή τη διαδικασία μπορεί στην πραγματικότητα να περιέχουν χρήσιμες πληροφορίες.

Μεγάλα Δεδομένα στη Ναυτιλία – Περιορισμοί και Προκλήσεις

- Η ποσότητα δεδομένων οδηγεί σε ζητήματα που αφορούν την αποθήκευση και μεταφορά δεδομένων μεταξύ πλοίου και ακτής.
- Υπάρχουν περιορισμοί στο μέγιστο διαθέσιμο εύρος ζώνης και το κόστος μετάδοσης μέσω δορυφορικής επικοινωνίας ανοικτής θάλασσας, τουλάχιστον σε ορισμένες περιοχές.
- Χρησιμοποιώντας ένα υποσύνολο του συνόλου δεδομένων μεγάλης κλίμακας, στατιστική ανάλυση για την τυποποίηση δεδομένων και τεχνικές μηχανικής μάθησης μπορεί να υποστηρίξουν αυτό το ζήτημα.
- Στην περίπτωση χρήσης δεδομένων που έχουν παραχθεί από μοντέλα ή προσομοιώσεις, η **αβεβαιότητα μοντέλου** είναι ένα από τα κύρια ζητήματα.
- Απαραίτητη είναι σε αυτές τις περιπτώσεις η αξιολόγηση – επαλήθευση του μοντέλου.

Μεγάλα Δεδομένα στη Ναυτιλία – Περιορισμοί και Προκλήσεις

- Μια άλλη σημαντική πρόκληση είναι η έλλειψη ειδικών στην μηχανική μεγάλων δεδομένων ειδικά στη ναυτιλιακή βιομηχανία – Η πλειοψηφία των ναυτιλιακών ηγετών/εταιρειών συμφωνεί ότι η σοβαρή έλλειψη δεξιοτήτων εμποδίζει τον κλάδο από την αποτελεσματική αξιοποίηση των Big Data.
- Σε όλους τους κλάδους, τα Big Data εγείρουν σοβαρά ηθικά και ζητήματα απορρήτου και την ανάγκη για ένα νομοθετικό πλαίσιο που θα υπαγορεύει τα όρια και το πλαίσιο χρήσης δεδομένων.
- Η DNV-GL θεωρεί ότι η καθιερωμένη διακυβέρνηση των δεδομένων είναι ένα από τα θεμελιώδη ζητήματα. Επισημαίνεται στο ότι, αν υπάρχουν διαφορετικά είδη δεδομένων ενδιαφερόμενοι που θα πρέπει να έχουν δικαιώματα ιδιοκτησίας μπορεί να διαφέρουν ενώ τα δικαιώματα ορισμένων δεδομένων δεν πρέπει μονοπωλούνται.

Μεγάλα Δεδομένα στη Ναυτιλία – Περιορισμοί και Προκλήσεις – Ασφάλεια Δεδομένων

- Οποιαδήποτε φυσική σύνδεση σε δίκτυα ή συστήματα μπορεί να είναι ευάλωτη σε εχθρικές επιθέσεις και αναφορές που αποστέλλονται από το πλοίο προς την η ακτή ή το αντίστροφο.
- *Cyber Security*
- *Data Privacy*
- *Data Misreporting*