



UNIVERSITY OF PIRAEUS
Department of Informatics
M.Sc. "Digital Culture, Smart Cities, IoT and Advanced Digital Technologies"

M2M ΕΠΙΚΟΙΝΩΝΙΕΣ

ΜΑΘΗΜΑ 2^ο

Μάριος Τυροβολάς, Ph.D. Student
tirovolas@kic.uoi.gr ή m.tyrovolas@uoi.gr

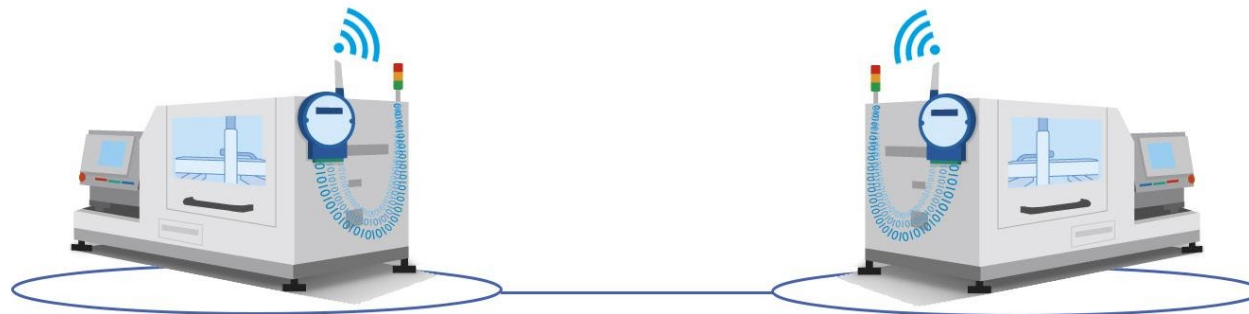
Πανεπιστήμιο Ιωαννίνων
Τμήμα Πληροφορικής και Τηλεπικοινωνιών

ΠΕΡΙΕΧΟΜΕΝΑ

- Σύνοψη προηγούμενου μαθήματος
- MQTT
- Αφότου συλλέξω τα δεδομένα;
- Πρακτικά παραδείγματα

ΣΥΝΟΨΗ ΠΡΟΗΓΟΥΜΕΝΟΥ ΜΑΘΗΜΑΤΟΣ

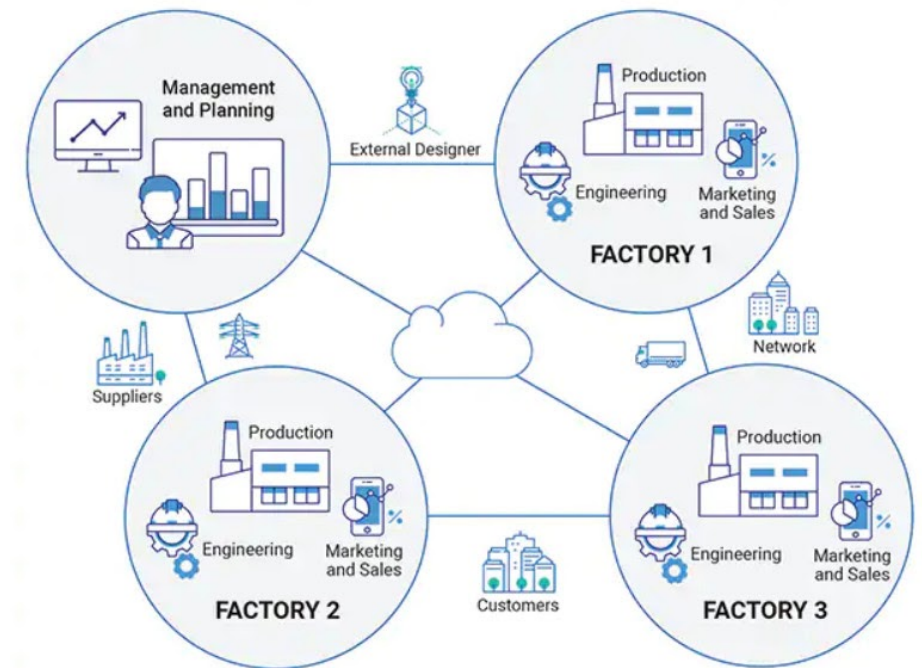
- Το Machine-to-Machine ή M2M δίκτυο αναφέρεται σε ένα δίκτυο συσκευών (μηχανών, αισθητήρων κ.λπ.) που ανταλλάσσουν δεδομένα μεταξύ τους χωρίς την ανθρώπινη παρέμβαση με σκοπό την αυτοματοποίηση της διαδικασίας.



ΣΥΝΟΨΗ ΠΡΟΗΓΟΥΜΕΝΟΥ ΜΑΘΗΜΑΤΟΣ

Το IoT/M2M φιλοδοξεί να προσφέρει:

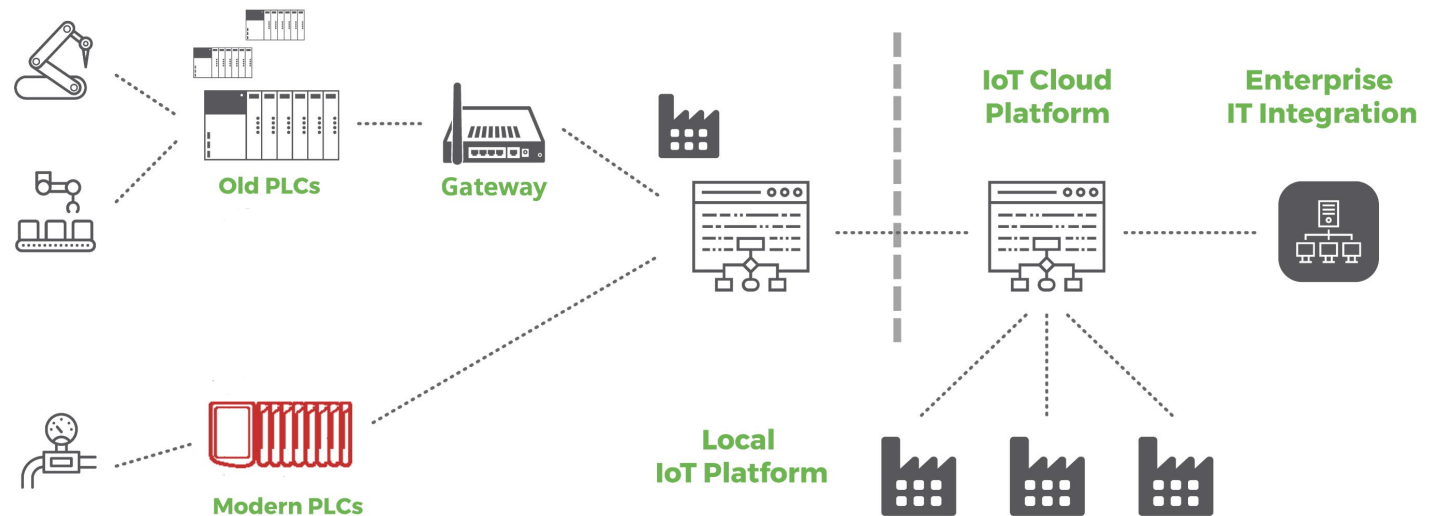
- Δυναμικό έλεγχο και υλοποίηση κατανεμημένων διαδικασιών στη βιομηχανία.
- Καλύτερη εποπτεία των συστημάτων παραγωγής με σκοπό την βελτιστοποίησή τους.
- Αύξηση της παραγωγικότητας
- Εξοικονόμηση χρόνου & Προϊόντα υψηλότερης ποιότητας



ΣΥΝΟΨΗ ΠΡΟΗΓΟΥΜΕΝΟΥ ΜΑΘΗΜΑΤΟΣ

Ωστόσο, οι βιομηχανίες διατηρούν και χρησιμοποιούν τον παλιό τους εξοπλισμό (ελεγκτές, μηχανές κ.ο.κ.) αλλά και τα παραδοσιακά συστήματα παραγωγής γιατί:

**Legacy
Modernization
for Manufacturing**

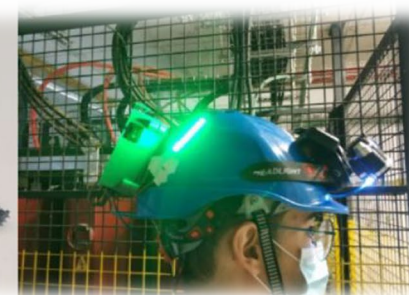
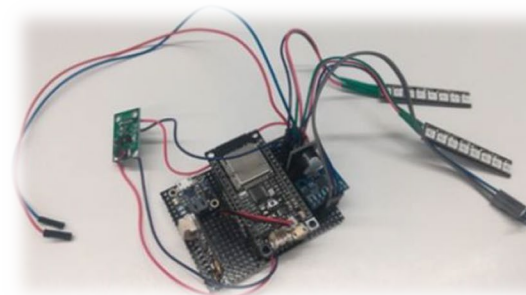
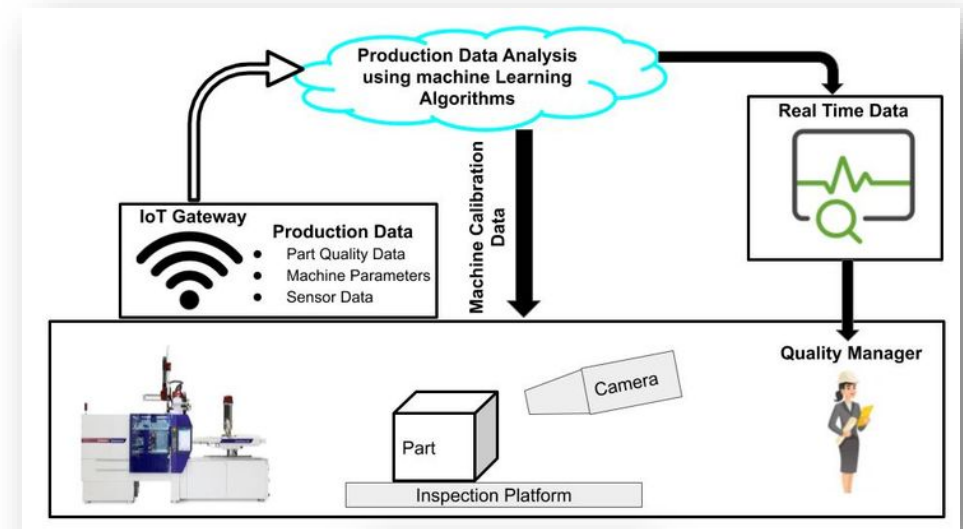
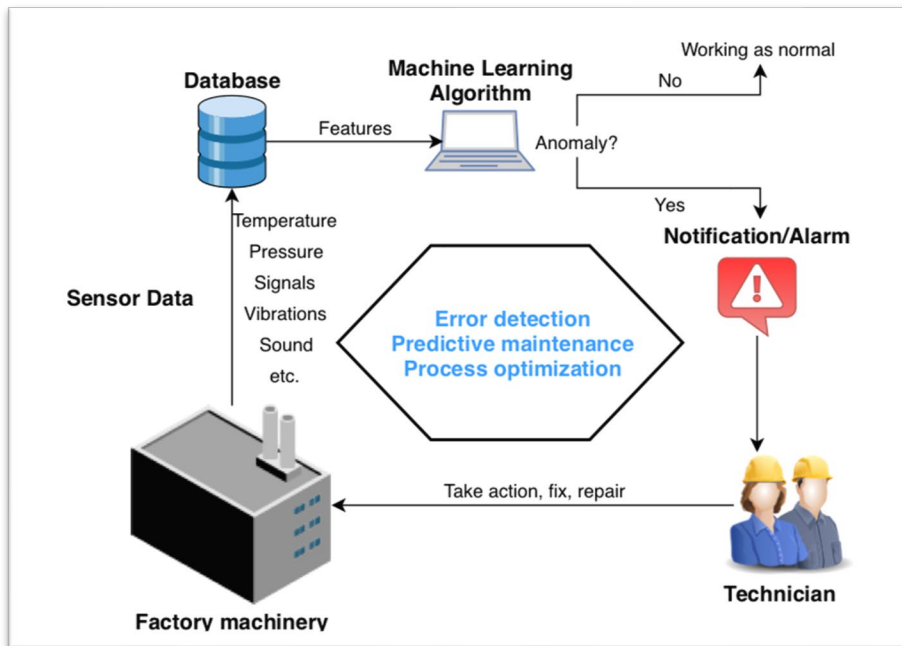


ΣΥΝΟΨΗ ΠΡΟΗΓΟΥΜΕΝΟΥ ΜΑΘΗΜΑΤΟΣ

Ευφυείς εφαρμογές αξιοποιώντας τις M2M επικοινωνίες:

- Παρακολούθηση σε πραγματικό χρόνο παραμέτρων του βιομηχανικού συστήματος και λήψη ειδοποιήσεων όταν εντοπίζονται ανωμαλίες (π.χ. υπερθέρμανση κινητήρα και υπερβολικοί κραδασμοί, διαρροή βαλβίδων/σωλήνων).
- Προγνωστική Συντήρηση (Predictive Maintenance)
- Έλεγχος ποιότητας (Quality Control)
- Μαζική εξατομίκευση (Mass Customization)
- Fail Soft/ Fail safe : Όταν μια μηχανή οδηγείται σε σφάλμα «ενημερώνει» το σύστημα παραγωγής με αποτέλεσμα αυτό να προσαρμόζεται είτε μειώνοντας τον ρυθμό παραγωγής είτε διακόπτοντας την λειτουργία του.
- Smart Personal Protection Equipment
- Παρακολούθηση της κατανάλωσης ενέργειας του εξοπλισμού και των συστημάτων για τον εντοπισμό των πηγών αποβλήτων.

ΣΥΝΟΨΗ ΠΡΟΗΓΟΥΜΕΝΟΥ ΜΑΘΗΜΑΤΟΣ

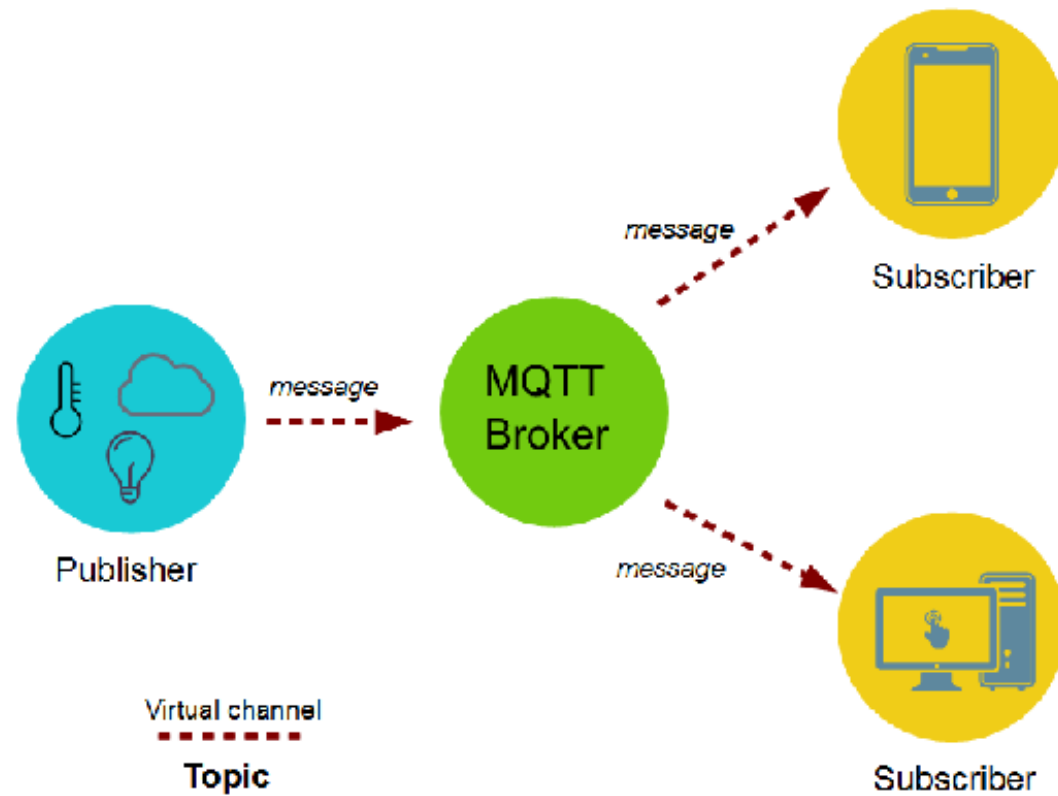


MQTT

- Υλοποιεί το μοντέλο **Publish/Subscribe** για την ανταλλαγή μηνυμάτων, όπου ένας κόμβος του δικτύου, ο **publisher**, στέλνει μηνύματα σε έναν άλλον κόμβο (ή περισσότερους), τον **subscriber**, με την χρήση ενός ενδιάμεσου διαχειριστή μηνυμάτων (**Broker**) ο οποίος φιλτράρει τα μηνύματα και τα παραδίδει στους ενδιαφερόμενους κόμβους. Οι publishers δεν γνωρίζουν ποιοί είναι οι subscribers αλλά και που βρίσκονται
- Αυτό το μοντέλο επικοινωνίας είναι ιδανικό για την διασύνδεση κατανεμημένων συστημάτων που είναι «χαλαρά» συνδεδεμένα και είναι συμβατό με τη λογική και πρακτική του IoT, όπου συσκευές ή αντικείμενα παράγουν γεγονότα που θέλουν να διαμοιραστούν.



MQTT



MQTT

- Το MQTT επιτρέπει τη δημιουργία θεματικών ενοτήτων (topics) για να απλουστεύσει την διαδικασία αποστολής μηνυμάτων και να επιτρέπει στους κόμβους να δημοσιεύουν δεδομένα σε αυτές.
- Τα θέματα χρησιμοποιούνται για να διακρίνει ο broker τον subscriber/subscribers που ενδιαφέρεται για τα μηνύματα.
- Οι κόμβοι μπορούν να εγγραφούν σε μια ενότητα και να ενημερώνονται με ασύγχρονες ειδοποιήσεις για την προσθήκη καινούργιων δεδομένων σε αυτές. Αυτός που αναλαμβάνει την ειδοποίηση είναι ο broker.
- Τα topic έχουν την μορφή ενός UTF-8 string και μπορούν να αποτελούνται από διαφορετικά επίπεδα τα οποία χωρίζονται με τον χαρακτήρα “/”.

MQTT

Ποιότητα Υπηρεσιών

Το MQTT υποστηρίζει 3 επίπεδα Quality of Service (QoS) κατά την αποστολή των μηνυμάτων

1. QoS 0 At most once

Σε αυτό το σημείο ο publisher στέλνει το μήνυμα στον Broker χωρίς να περιμένει απάντηση

2. QoS 1 At least once

Μόλις ληφθεί το μήνυμα από τον Broker, το διανέμει στους subscribers ενώ παράλληλα προωθεί την επιβεβαίωση παραλαβής στον publisher (PUBACK). Αν ο publisher δεν λάβει το μήνυμα θα επιχειρήσει να ξαναστείλει το πακέτο το οποίο έχει αποθηκεύσει. Για να γίνει η συσχέτιση του μηνύματος που στάλθηκε με το μήνυμα που παραλήφθηκε χρησιμοποιούνται αναγνωριστικά στα μηνύματα.

3. QoS 2 Exactly Once

Η παράδοση του μηνύματος είναι εγγυημένη χωρίς πιθανότητα επανάληψης αποστολής. Ο publisher στέλνει μήνυμα και ο broker

MQTT



MQTT Client

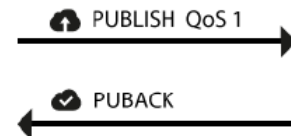


MQTT Broker

PUBLISH: Η δημοσίευση ενός καινούριου μηνύματος



MQTT Client

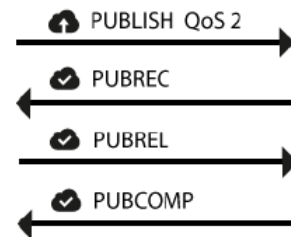


MQTT Broker

PUBACK: Απάντηση σε ένα μήνυμα PUBLISH με QoS 1



MQTT Client



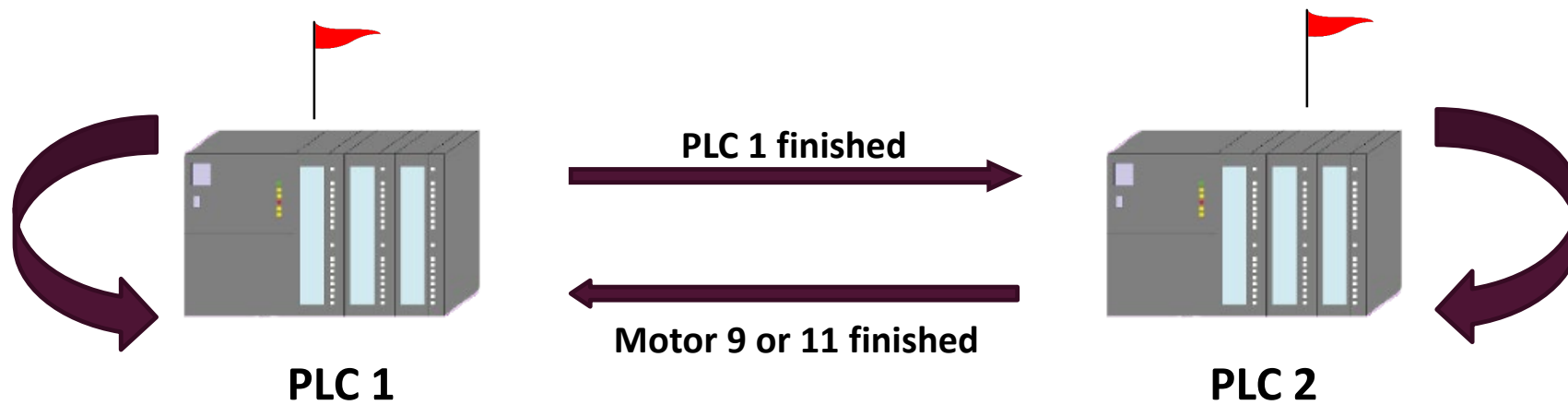
MQTT Broker

PUBREC: Το πρώτο κομμάτι μιας σειράς μηνυμάτων QoS 2

PUBREL: Το δεύτερο κομμάτι μιας σειράς μηνυμάτων QoS 2

PUBCOMP: Το τελευταίο κομμάτι μιας σειράς μηνυμάτων QoS 2

MQTT Εφαρμογή



ΠΡΑΚΤΙΚΑ ΠΑΡΑΔΕΙΓΜΑΤΑ

- MQTT (<http://www.hivemq.com/demos/websocket-client/>) & MQTIZER (Android)

ΑΦΟΤΟΥ ΣΥΛΛΕΞΩ ΤΑ ΔΕΔΟΜΕΝΑ;

ΑΦΟΤΟΥ ΣΥΛΛΕΞΩ ΤΑ ΔΕΔΟΜΕΝΑ;

- Ο μεγάλος όγκος δεδομένων που συλλέγεται πλέον από τους αισθητήρες, περιέχει πολύ σημαντικές πληροφορίες για το σύστημα μας. Ωστόσο μέχρι σήμερα, παραπάνω από το 75% των δεδομένων παραμένουν αναξιοποίητα.
- Πως μπορούμε όμως να αξιοποιήσουμε αυτά τα δεδομένα για την υλοποίηση έξυπνων εφαρμογών?

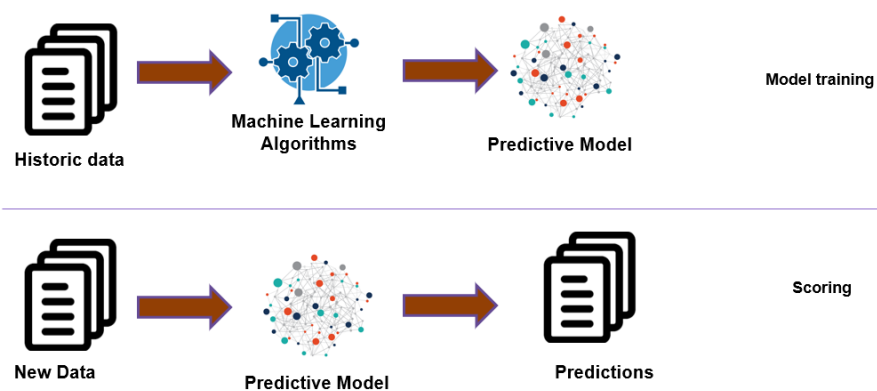
Μέσω της **Μηχανικής Μάθησης (Machine Learning)**!

ΜΗΧΑΝΙΚΗ ΜΑΘΗΣΗ

Τι είναι η Μηχανική Μάθηση;

Η μηχανική μάθηση είναι υποπεδίο της τεχνητής νοημοσύνης **διερευνά τη μελέτη και την κατασκευή συστημάτων (αλγορίθμων) που μπορούν να μαθαίνουν από δεδομένα και να κάνουν προβλέψεις σχετικά με αυτά.**

Η "πρόβλεψη" ενός αλγορίθμου μηχανικής μάθησης αναφέρεται στην τιμή της εξόδου του για νέα δεδομένα εισόδου αφού έχει εκπαιδευτεί. Οι προβλέψεις



Με βάση τα δεδομένα που έχω εκπαιδευτεί, η τιμή της εξόδου για αυτά τα νέα δεδομένα εισόδου είναι :....

ΕΙΔΗ ΠΡΟΒΛΗΜΑΤΩΝ ΜΗΧΑΝΙΚΗΣ ΜΑΘΗΣΗΣ

Χαρακτηριστικά προβλήματα που μπορούμε να λύσουμε με την μηχανική μάθηση;

- Ταξινόμησης (Classification)
 - Παλινδρόμησης (Regression)
 - Ομαδοποίησης (Clustering)
- Επιβλεπόμενη μάθηση
(Supervised Learning)
- Μη Επιβλεπόμενη μάθηση
(Unsupervised Learning)

ΠΡΟΒΛΗΜΑ ΤΑΞΙΝΟΜΗΣΗΣ

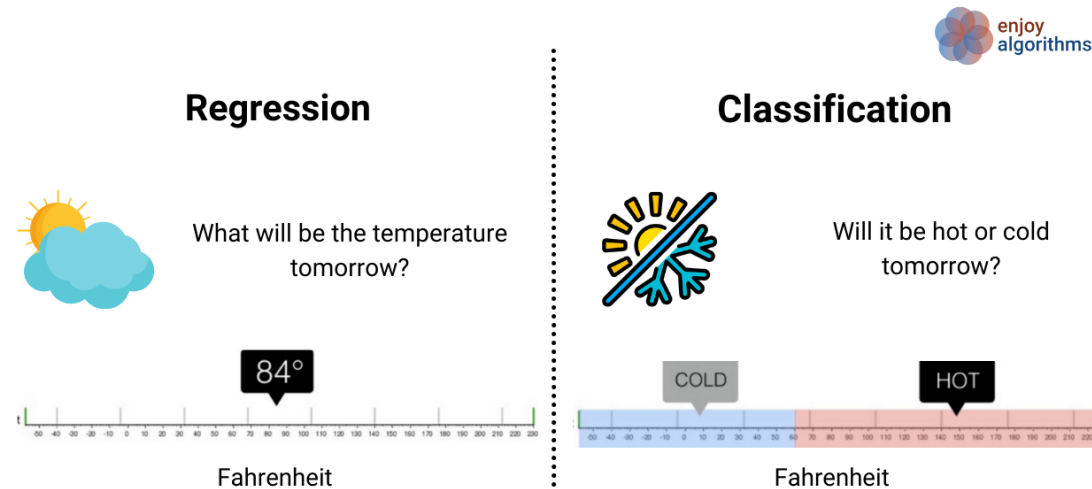
- Ο αλγόριθμος (ή μοντέλο) μηχανικής μάθησης κατηγοριοποιεί τα δεδομένα εισόδου σε συγκεκριμένες και προκαθορισμένες ομάδες, γνωστές και ως **κλάσεις**. Συνεπώς, η έξοδος παίρνει **διακριτές τιμές**
- 3 είδη προβλημάτων ταξινόμησης
 1. Binary Classification Problem (Spam or No Spam Email)
 2. Multi-class Classification Problem (Dog breeds)
 3. Multi Label Classification Problem (Film Genres)

Ερώτηση: Ποιες από τις προηγούμενες εφαρμογές αφορούν πρόβλημα ταξινόμησης;

ΠΡΟΒΛΗΜΑ ΠΑΛΙΝΔΡΟΜΗΣΗΣ

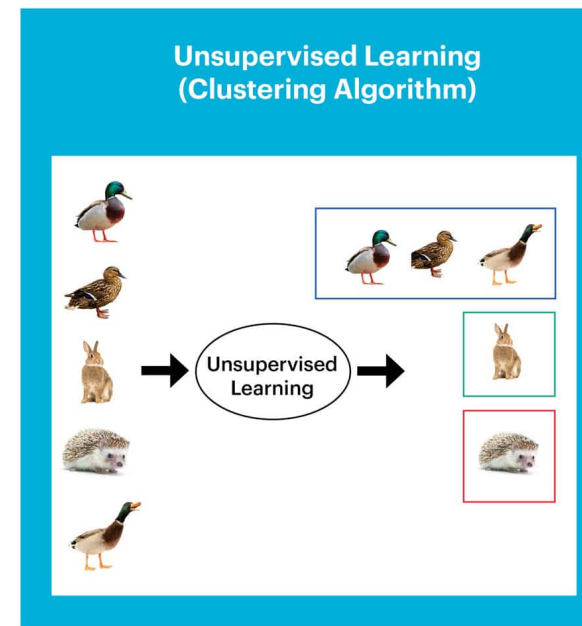
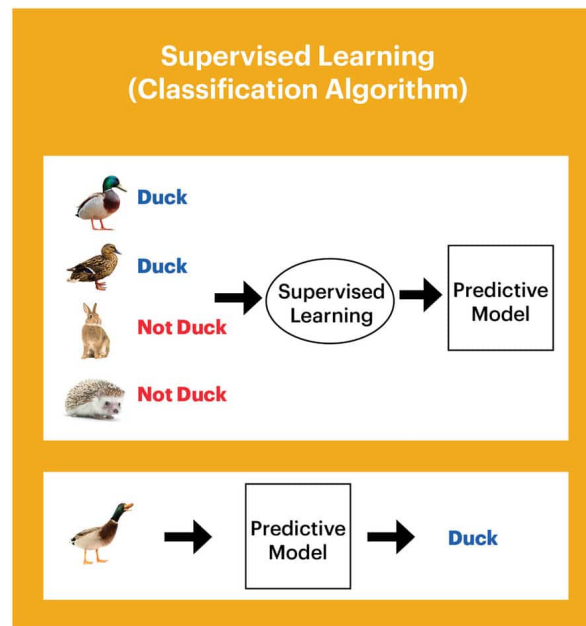
Το πρόβλημα της παλινδρόμησης αφορά την πρόβλεψη μιας **συνεχούς ποσότητας** (π.χ. η χρηματική τιμή ενός αυτοκινήτου δεδομένου του αριθμού των πορτών, την ιπποδύναμη της μηχανής, την χρονολογία κατασκευής του κ.ο.κ)

Ερώτηση: Ποιες από τις προηγούμενες εφαρμογές αφορούν πρόβλημα πρόγνωσης;

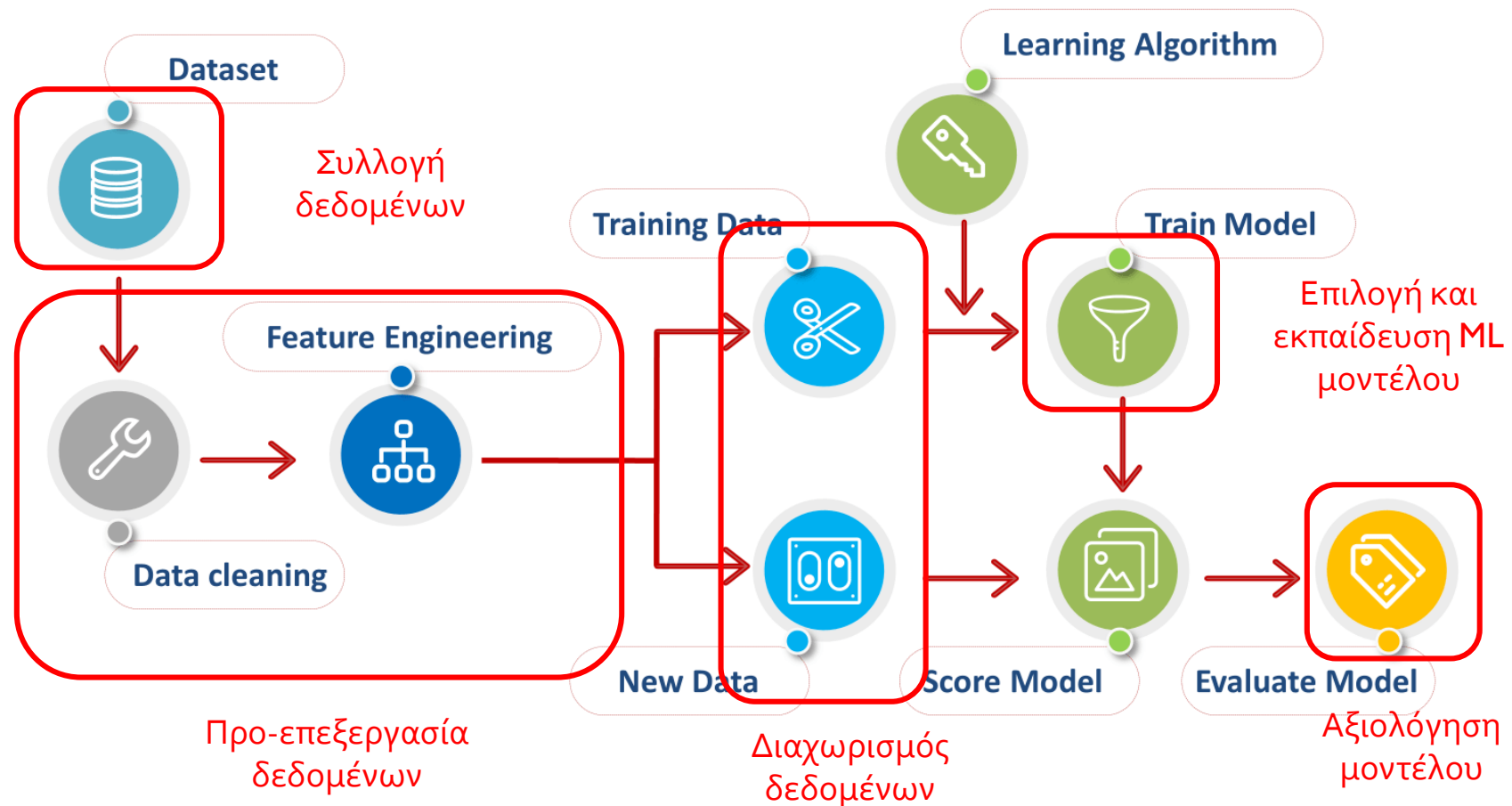


ΠΡΟΒΛΗΜΑ ΟΜΑΔΟΠΟΙΗΣΗΣ

- Στο πρόβλημα της ομαδοποίησης (clustering) ένα σύνολο δεδομένων πρόκειται να χωριστεί σε ομάδες. Σε αντίθεση με την ταξινόμηση, **οι ομάδες δεν είναι γνωστές** εκ των προτέρων, καθιστώντας αυτόν τον διαχωρισμό τυπικό πρόβλημα μη επιβλεπόμενης μάθησης.



ΣΤΑΔΙΑ ΜΗΧΑΝΙΚΗΣ ΜΑΘΗΣΗΣ



ΣΥΛΛΟΓΗ ΔΕΔΟΜΕΝΩΝ

- Τα δεδομένα από τους αισθητήρες συλλέγονται σε ένα πίνακα δεδομένων που ονομάζεται **dataset**.
- Κάθε στήλη αυτού του πίνακα (λέγεται **attribute** ή **feature**) αντιστοιχεί σε μία παράμετρο (στοιχείο) του συστήματος (π.χ θερμοκρασία, επίπεδο δόνησης μηχανής κ.ο.κ)
- Υπάρχουν **2 ειδών datasets**, όπου το κάθε ένα αντιστοιχεί στο είδος του προβλήματος που θέλουμε να λύσουμε (πρόβλημα επιβλεπόμενης ή μη επιβλεπόμενης μάθησης). Το 1^ο είδος λέγεται **labeled dataset** και το 2^ο είδος **unlabeled dataset**.

Labeled dataset —————> Πρόβλημα ταξινόμησης ή παλινδρόμησης

Unlabeled dataset —————> Πρόβλημα ομαδοποίησης

LABELLED DATASET

Έστω θέλουμε να χτίσουμε ένα μοντέλο μηχανικής μάθησης το οποίο θα δέχεται σαν είσοδο τα δεδομένα της βιοψίας διαφόρων ασθενών και θα προβλέπει αν ο εξεταζόμενος όγκος είναι καλοήθης (benign) ή κακοήθης (malignant)

Ερώτηση: Τί είδους πρόβλημα είναι αυτό;

ID	Clump	UnifSize	UnifShape	MargAdh	SingEpiSize	BareNuc	BlandChrom	NormNucl	Mit	Class
1000025	5	1	1	1	2	1	3	1	1	benign
1002945	5	4	4	5	7	10	3	2	1	benign
1015425	3	1	1	1	2	2	3	1	1	malignant
1016277	6	8	8	1	3	4	3	7	1	benign
1017023	4	1	1	3	2	1	3	1	1	benign
1017122	8	10	10	8	7	10		7	1	malignant
1018099	1	1	1	1	2	10	3	1	1	benign
1018561	2	1	2	H	2	1	3	1	1	benign
1033078	2	1	1	1	2	1	1	1	5	benign
1033078	4	2	1	1	2	1	2	1	1	benign

labels

Πέρα από τα δεδομένα εισόδου, σε κάθε καταγραφή έχουμε και την τιμή της εξόδου (target variable)

LABELED DATASET

Το ίδιο ισχύει και στα προβλήματα παλινδρόμησης:

Δεδομένα εισόδου			Έξοδος
<i>Fuel type</i>	<i>The number of doors</i>	<i>Engine size (61 – 326)</i>	Price
1	4	70	11000
1	2	120	17000
2	2	310	41000

UNLABELED DATASET

- Η διαφορά με το labeled dataset είναι ότι περιέχονται μόνο τα δεδομένα εισόδου χωρίς να είναι γνωστή η έξοδος για κάθε καταγραφή

Customer Id	Age	Edu	Years Employed	Income	Card Debt	Other Debt	Address	DebtIncomeRatio
1	41	2	6	19	0.124	1.073	NBA001	6.3
2	47	1	26	100	4.582	8.218	NBA021	12.8
3	33	2	10	57	6.111	5.802	NBA013	20.9
4	29	2	4	19	0.681	0.516	NBA009	6.3
5	47	1	31	253	9.308	8.908	NBA008	7.2
6	40	1	23	81	0.998	7.831	NBA016	10.9
7	38	2	4	56	0.442	0.454	NBA013	1.6
8	42	3	0	64	0.279	3.945	NBA009	6.6
9	26	1	5	18	0.575	2.215	NBA006	15.5
10	47	3	23	115	0.653	3.947	NBA011	4
11	44	3	8	88	0.285	5.083	NBA010	6.1
12	34	2	9	40	0.374	0.266	NBA003	1.6

unlabeled

ΠΡΟ-ΕΠΕΞΕΡΓΑΣΙΑ ΔΕΔΟΜΕΝΩΝ

Αφότου έχουμε συλλέξει τα δεδομένα, ξεκινά η προ-επεξεργασία:

1. Αφαίρεση θορύβου.
2. Μετατροπή όλων των δεδομένων σε αριθμητικές τιμές καθώς τα machine learning μοντέλα έχουν καλύτερη απόδοση όταν χρησιμοποιούν αριθμητικά δεδομένα.
3. Αντικατάσταση δεδομένων που λείπουν με μια εκτίμηση (συνήθως με την μέση τιμή του αντίστοιχου feature).
4. Επιλογή των features που συμβάλλουν περισσότερο στην πρόβλεψη της εξόδου (feature selection).
5. Κανονικοποίηση των δεδομένων ούτως ώστε κάθε feature να συμβάλλει προσεγγιστικά το ίδιο στην πρόβλεψη της εξόδου. Αρκετοί αλγόριθμοι μηχανικής μάθησης είναι ευαίσθητοι στο εύρος των τιμών του κάθε feature .

ΔΙΑΧΩΡΙΣΜΟΣ ΔΕΔΟΜΕΝΩΝ

Αφότου έχουμε προ-επεξεργαστεί το dataset και είναι έτοιμο για χρήση, χωρίζεται κυρίως σε 2 ομάδες:

- **Training Dataset:** Χρησιμοποιείται για την εκπαίδευση του μοντέλου
- **Test Dataset:** Χρησιμοποιείται για την αξιολόγηση του εκπαιδευμένου μοντέλου

ΕΠΙΛΟΓΗ ΚΑΙ ΕΚΠΑΙΔΕΥΣΗ ΜΟΝΤΕΛΟΥ

Αφότου έχουμε χωρίσει τα δεδομένα, **επιλέγουμε** τον αλγόριθμο μηχανικής μάθησης που θα χρησιμοποιήσουμε και τον εκπαιδεύουμε χρησιμοποιώντας το `training dataset`.

- Για κάθε είδος προβλήματος υπάρχουν διαφορετικοί αλγόριθμοι που μπορούν να χρησιμοποιηθούν.
- Η επιλογή του καταλληλότερου αλγορίθμου γίνεται βάση των μετρικών αξιολόγησης.

Κατά την **εκπαίδευση** του αλγορίθμου, σκοπός μας είναι να βρούμε τις βέλτιστες τιμές των παραμέτρων του μοντέλου ούτως ώστε για κάθε δοσμένο σύνολο των δεδομένων εισόδου, η έξοδος του μοντέλου να είναι όσο το δυνατόν πιο κοντά στην επιθυμητή.

ΑΞΙΟΛΟΓΗΣΗ ΤΟΥ ΜΟΝΤΕΛΟΥ

Αφότου ολοκληρωθεί η εκπαίδευση του μοντέλου, νέα δεδομένα (test set) χρησιμοποιούνται προκειμένου να αξιολογήσουμε την απόδοση του μοντέλου (για κάθε παράδειγμα δεδομένων εισόδου του test set, το μοντέλο προβλέπει την τιμή της εξόδου και στη συνέχεια χρησιμοποιεί τις κατάλληλες μετρικές αξιολόγησης).

Ανάλογα με το είδος του προβλήματος (ταξινόμησης ή πρόβλεψης), χρησιμοποιούμε και άλλες μετρικές αξιολόγησης.

Metric	Formula	Evaluation Focus
Accuracy	$ACC = \frac{TP+TN}{TP+TN+FP+FN}$	Overall effectiveness of a classifier
Error rate	$ERR = \frac{FP+FN}{TP+TN+FP+FN}$	Classification error
Precision	$PRC = \frac{TP}{TP+FP}$	Class agreement of the data labels with the positive labels given by the classifier
Sensitivity	$SNS = \frac{TP}{TP+FN}$	Effectiveness of a classifier to identify positive labels
Specificity	$SPC = \frac{TN}{TN+FP}$	How effectively a classifier identifies negative labels
ROC	$ROC = \frac{\sqrt{SNS^2+SPC^2}}{\sqrt{2}}$	Combined metric based on the Receiver Operating Characteristic (ROC) space [53]
F_1 score	$F_1 = 2 \frac{PRC \cdot SNS}{PRC + SNS}$	Combination of precision (PRC) and sensitivity (SNS) in a single metric
Geometric Mean	$GM = \sqrt{SNS \cdot SPC}$	Combination of sensitivity (SNS) and specificity (SPC) in a single metric

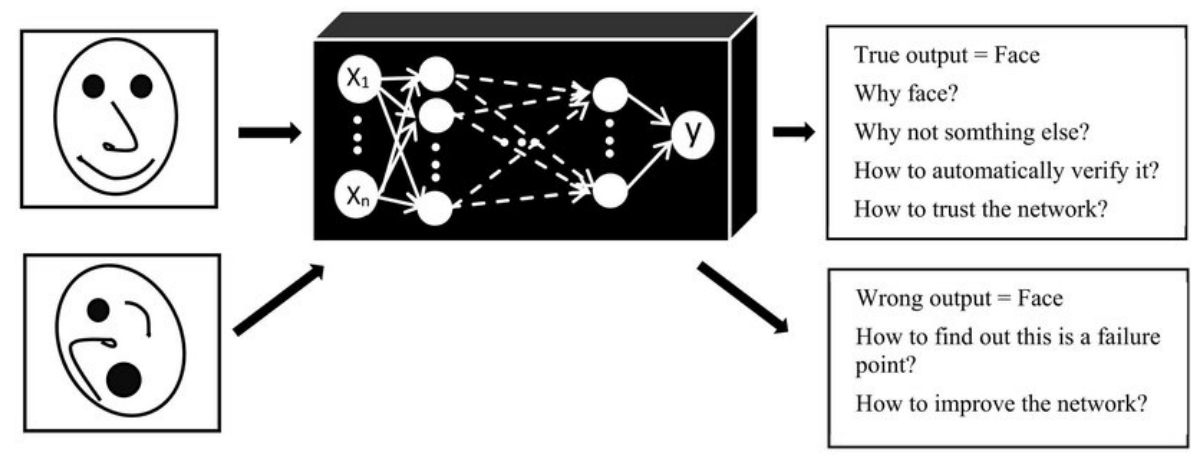
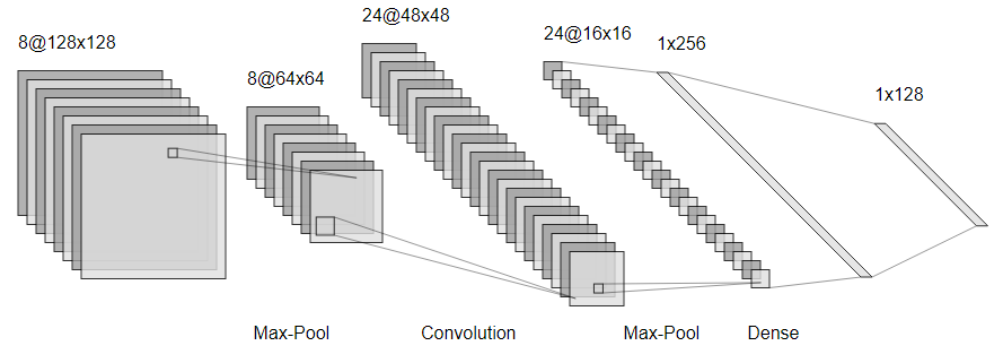
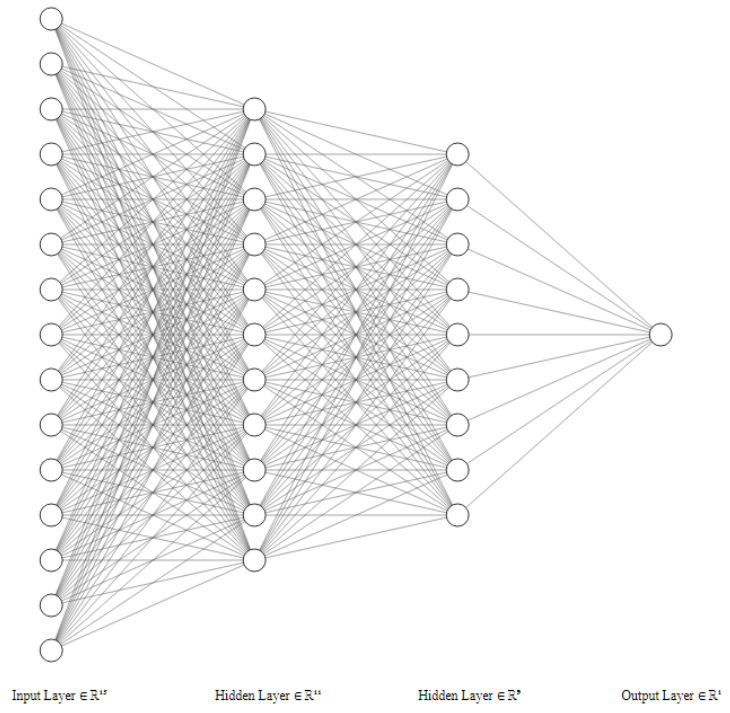
Mean squared error	$MSE = \frac{1}{n} \sum_{t=1}^n e_t^2$
Root mean squared error	$RMSE = \sqrt{\frac{1}{n} \sum_{t=1}^n e_t^2}$
Mean absolute error	$MAE = \frac{1}{n} \sum_{t=1}^n e_t $
Mean absolute percentage error	$MAPE = \frac{100\%}{n} \sum_{t=1}^n \left \frac{e_t}{y_t} \right $

ΠΡΑΚΤΙΚΑ ΠΑΡΑΔΕΙΓΜΑΤΑ

- Quality Control Example (app.ximilar.com)
- Regression Problem (**Jupyter Notebook**)

ΟΡΙΣΜΕΝΑ ΜΕΙΟΝΕΚΤΗΜΑΤΑ ΤΗΣ ΜΗΧΑΝΙΚΗΣ ΜΑΘΗΣΗΣ

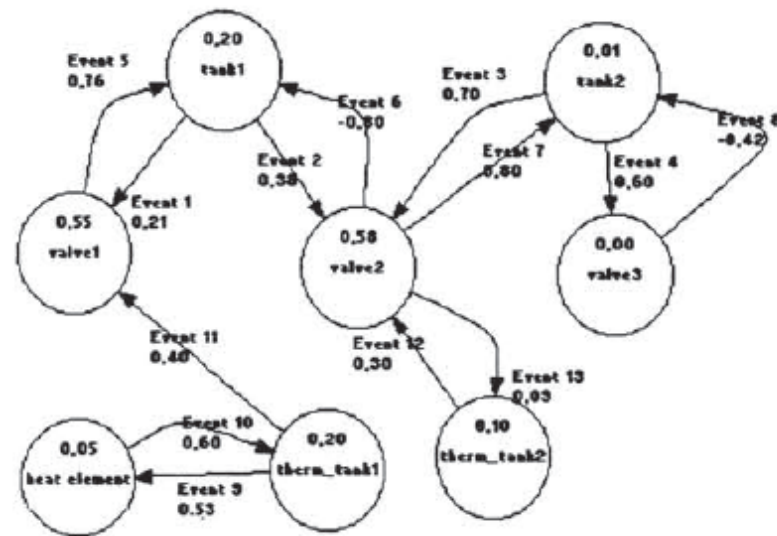
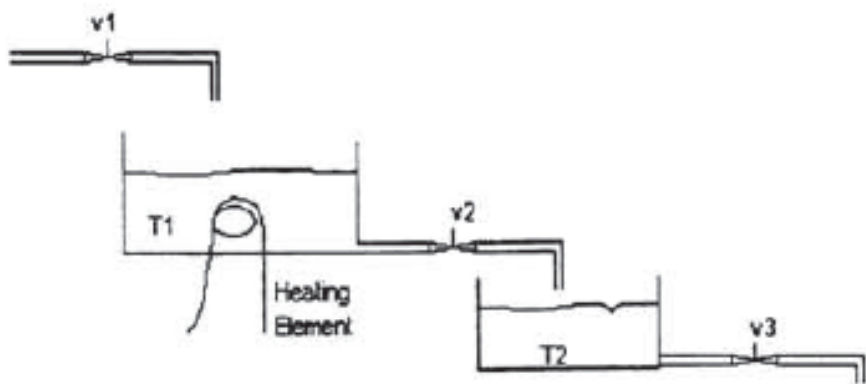
- Όταν τα δεδομένα που έχουμε συλλέξει δεν είναι αρκετά, το μοντέλο έχει μειωμένη απόδοση.
- Πρόκειται για πολύπλοκα μοντέλα (όσον αφορά τη δομή) μέσω των οποίων είναι δύσκολο να επεξηγηθεί πως και για ποιο λόγο τα συγκεκριμένα δεδομένα εισόδου οδήγησαν σε αυτή την πρόβλεψη (π.χ. ιατρικές εφαρμογές)



ΠΩΣ ΜΠΟΡΟΥΝ ΝΑ ΑΝΤΙΜΕΤΩΠΙΣΤΟΥΝ ;

Μεθοδολογίες εύκαμπτης πληροφορικής (Soft Computing) —> Ασαφή Γνωστικά Δίκτυα (Fuzzy Cognitive Maps)

- Χρησιμοποιούνται για την μοντελοποίηση πολύπλοκων συστημάτων αλλά και για προβλήματα λήψης απόφασης (decision-making processes) {Τα προβλήματα τεχνητής νοημοσύνης αποτελούν προβλήματα λήψης απόφασης}
- Διαισθητική (γραφική) αναπαράσταση των σχέσεων αιτίου-αποτελέσματος μεταξύ των στοιχείων ενός συστήματος. Κάθε κόμβος αποτελεί μια μεταβλητή του συστήματος ενώ τα βάρη εκφράζουν τον βαθμό αιτιότητας μεταξύ τους.





ΑΝΑΛΟΓΕΣ ΕΦΑΡΜΟΓΕΣ ΣΤΙΣ ΕΞΥΠΝΕΣ ΠΟΛΕΙΣ;

Ευχαριστώ
πολύ!