

Βασικές Έννοιες στην εποχή της AI / ML

**ΝΕΟΙ ΟΡΙΖΟΝΤΕΣ ΤΗΣ
ΘΕΩΡΙΑΣ ΤΗΛΕΠΙΚΟΙΝΩΝΙΑΚΗΣ ΚΙΝΗΣΕΩΣ**

Βασικοί όροι της Μηχανικής Μάθησης

- **Μοτίβο (pattern):** Κάτι που συμβαίνει συχνά (επανεμφανιζόμενη δομή, σχέση ή κανονικότητα) μέσα σε δεδομένα, το οποίο μπορεί να ανακαλύψει ώστε να κάνει προβλέψεις (να πάρει αποφάσεις). Π.χ. μοτίβο θερμοκρασίας με πωλήσεις παγωτών.
- **Δεδομένα (dataset):** Απαιτείται αφθονία δεδομένων (αριθμοί, εικόνες, κείμενα, ήχοι κλπ), ώστε οι αλγόριθμοι Μηχανικής Μάθησης να μπορούν να βρίσκουν μοτίβα.
- **Αλγόριθμος:** Σύνολο κανόνων ή μαθηματικών μεθόδων που μαθαίνουν από δεδομένα, ώστε να αναγνωρίζουν μοτίβα και να προβλέπουν/αποφασίζουν χωρίς ξεχωριστό προγραμματισμό για κάθε περίπτωση.
- **Μοντέλο:** Μαθηματική αναπαράσταση των μαθησιακών μοτίβων (αποτέλεσμα εκπαίδευσης), που χρησιμοποιείται για προβλέψεις.
- **Εκπαίδευση:** Η διαδικασία τροφοδοσίας δεδομένων σε έναν αλγόριθμο για την κατασκευή του μοντέλου.
- **Πρόβλεψη/Απόφαση:** Χρήση του εκπαιδευμένου μοντέλου (βάσει των δεδομένων) για λήψη αποφάσεων ή προβλέψεων σε νέα δεδομένα.

Binary Confusion Matrix (Πίνακας Σύγχυσης)

- Χρησιμοποιείται στην μέτρηση της απόδοσης ενός μοντέλου ταξινόμησης (machine learning).
- Συγκρίνει τις προβλέψεις του μοντέλου με την πραγματικότητα και δείχνει πού το μοντέλο ήταν σωστό ή λάθος.
- Αναλύει τις προβλέψεις σε τέσσερις κατηγορίες (2 σωστές – 2 λάθος): (TP, TN) – (FP ψευδής συναγερμός, FN).

Confusion Matrix (2 × 2)

π ρ ό β λ ε ψ η	Confusion Matrix	Πραγματικότητα	
		Θετική (P)	Αρνητική (N)
Θετική (P)		TP	FP σφάλμα Τύπου I
Αρνητική (N)		FN σφάλμα Τύπου II	TN

Π ρ α γ μ α τ ι κ ό τ η τ α	Confusion Matrix	Πρόβλεψη	
		Θετική (P)	Αρνητική (N)
Θετική (P)		TP	FN σφάλμα Τύπου II
Αρνητική (N)		FP σφάλμα Τύπου I	TN

- TP: True Positive, TN: True Negative
- FP: False Positive, FN: False Negative

Basic Metrics (Βασικοί Δείκτες)

- **Precision** = $\frac{TP}{TP+FP}$ = Ακρίβεια

Πόσες θετικές προβλέψεις ήταν σωστές.

- **True Positive rate = Recall** = $\frac{TP}{TP+FN}$ = Ευαισθησία

Ποσοστό των πραγματικά θετικών αποτελεσμάτων που εντοπίστηκαν σωστά (ως θετικά).

- **False Positive rate** = $\frac{FP}{FP+TN}$ = False alarm Prob.

Ποσοστό των πραγματικών αρνητικών αποτελεσμάτων που εντοπίστηκαν εσφαλμένα ως θετικά.

- **True Negative rate** = $\frac{TN}{TN+FP}$ = Εξειδίκευση

Ποσοστό των πραγματικά αρνητικών αποτελεσμάτων που εντοπίστηκαν σωστά ως αρνητικά.

- **ACCURACY** = $\frac{TP+TN}{TP+TN+FB+FN}$ = Μέση Ακρίβεια

Ποσοστό των σωστών προβλέψεων.

- **F1 score** = $2 \frac{Precision \times Recall}{Precision + Recall}$ = Δείκτης ακρίβειας

Αρμονικός μέσος των τιμών Precision και Recall.

ΠΟΤΕ ΧΡΗΣΙΜΟΠΟΙΕΙΤΑΙ Ο ΚΑΘΕ ΔΕΙΚΤΗΣ

- **Precision:** Όταν οι θετικές προβλέψεις είναι σημαντικό να είναι σωστές.
- **Recall:** Όταν οι λανθασμένα αρνητικές προβλέψεις κοστίζουν περισσότερο από τις λανθασμένα θετικές. Π.χ. όταν η λανθασμένη γνωμάτευση απουσίας νόσου κοστίζει περισσότερο από λανθασμένη γνωμάτευση ύπαρξης νόσου.
- **FP rate:** Όταν οι λανθασμένα θετικές προβλέψεις κοστίζουν περισσότερο από τις λανθασμένα αρνητικές.
- **TN rate:** Όταν θέλουμε να είμαστε σίγουροι ότι κάτι δεν συμβαίνει (π.χ. νόσος, false alarm).
- **Accuracy:** Όταν τα σύνολα δεδομένων είναι **ισορροπημένα** ως πρόχειρος δείκτης προόδου/ σύγκλισης εκπαίδευσης μοντέλου.
- **F1 score:** Όταν υπάρχει ανισορροπία στα σύνολα δεδομένων (στην κατανομή των κατηγοριών), ή όταν το κόστος των ψευδώς θετικών (FP) και των ψευδώς αρνητικών (FN) προβλέψεων δεν είναι ίσο.

Confusion Matrix με πολλαπλές κατηγοριοποιήσεις

Confusion Matrix (3 × 3)

Πρόβλεψη	Confusion Matrix	Πραγματικότητα		
		Μήλο	Πορτοκάλι	Ροδάκινο
	Μήλο	TP	FN	FN
	Πορτοκάλι	FN	TP	FN
	Ροδάκινο	FN	FN	TP

Πραγματικότητα	Confusion Matrix	Πρόβλεψη		
		Μήλο	Πορτοκάλι	Ροδάκινο
	Μήλο	TP	FN	FN
	Πορτοκάλι	FN	TP	FN
	Ροδάκινο	FN	FN	TP

Στον πίνακα με πολλαπλές κατηγοριοποιήσεις εμφανίζονται μόνο οι όροι (μετρήσεις) TP και FN, αλλά όλοι οι όροι (TP, TN, FP and FN) ισχύουν όπως έχουν ορισθεί.

Παράδειγμα Δυαδικού Πίνακα Σύγχυσης

- Ελήφθησαν 100 emails, εκ των οποίων 40 ήταν spam και το μοντέλο σωστά αναγνώρισε 30 ως spam (TP).
- Ελήφθησαν 100 emails, εκ των οποίων 60 ΔΕΝ ήταν spam και το μοντέλο σωστά αναγνώρισε 50 ως NOT spam (TN).
- Ελήφθησαν 100 emails, από τα οποία το μοντέλο εσφαλμένα αναγνώρισε 10 κανονικά emails ως spam (FP).
- Ελήφθησαν 100 emails, από τα οποία το μοντέλο εσφαλμένα αναγνώρισε 5 ως NOT spam (ενώ ήταν spam) (FN).

TP = 30, TN = 50, FP = 10, FN = 5

π ρ ό β λ ε ψ η	Confusion Matrix	Πραγματικότητα	
		Spam	NOT Spam
	Spam	30	5
	NOT Spam	10	50

π ρ α γ μ α τ ι κ ό τ η τ α	Confusion Matrix	Πρόβλεψη	
		Email	Spam
	Email	30	10
	Spam	10	50

$$\text{Precision} = \frac{TP}{TP+FP} = \frac{30}{30+10} = 75\%, \quad \text{Recall} = \frac{TP}{TP+FN} = \frac{30}{30+5} = 85,7\%, \quad \text{FP rate} = \text{False Alarm} = \frac{10}{10+50} = 16,67\%$$

$$\text{TN rate} = \frac{TN}{TN+FP} = \frac{50}{50+10} = 83,33\%, \quad \text{ACCURACY} = \frac{TP+TN}{TP+TN+FP+FN} = \frac{30+50}{30+50+10+5} = 84,21\%, \quad \text{F1} = 2 \frac{0,75 \times 0,857}{0,75+0,857} = 80\%$$