KDD & Data Mining

Η διαφορά μεταξύ Ανακάλυψης Γνώσης (KDD) και Εξόρυξης Δεδομένων
Πληροφοριακά Συστήματα - Γ' ΕΠΑΛ

Η Βασική Διαφορά

Ανακάλυψη Γνώσης από Βάσεις Δεδομένων (KDD):

Είναι μια συνολική, ολοκληρωμένη διαδικασία μετατροπής των ακατέργαστων δεδομένων σε χρήσιμη και εφαρμόσιμη γνώση.

Εξόρυξη Δεδομένων (Data Mining):

Είναι απλώς ένα συγκεκριμένο, αλλά το πιο κρίσιμο στάδιο μέσα στη διαδικασία KDD.

Με απλά λόγια:

Το KDD είναι το όλο έργο και το Data Mining είναι το εργαλείο/βήμα που κάνει τη «βαριά» δουλειά της ανάλυσης.

Στάδια της Διαδικασίας KDD

1. Επιλογή (Selection): Επιλογή των κατάλληλων δεδομένων από τη βάση.

2. Προεπεξεργασία (Preprocessing): Καθαρισμός δεδομένων (αφαίρεση θορύβου, διόρθωση λαθών).

3. Μετασχηματισμός (Transformation): Μετατροπή των δεδομένων σε κατάλληλη μορφή για ανάλυση.

4. Εξόρυξη Δεδομένων (Data Mining): Εφαρμογή έξυπνων αλγορίθμων για την εύρεση μοτίβων.

5. Ερμηνεία / Αξιολόγηση (Interpretation / Evaluation): Κατανόηση των μοτίβων και μετατροπή τους σε χρήσιμη γνώση.

Τι ακριβώς είναι το Data Mining;

Το 4ο και πιο ουσιαστικό στάδιο του KDD.

Φάση εφαρμογής έξυπνων αλγορίθμων και στατιστικών μεθόδων πάνω στα προετοιμασμένα δεδομένα.

Κύριοι Στόχοι:

- Εντοπισμός κρυμμένων μοτίβων (patterns).
- Εύρεση άγνωστων συσχετίσεων μεταξύ των δεδομένων.
- Πρόβλεψη μελλοντικών τάσεων για τη λήψη αποφάσεων.

Σύγκριση: KDD vs Data Mining

Χαρακτηριστικό	Ανακάλυψη Γνώσης (KDD)	Εξόρυξη Δεδομένων (Data Mining)
Πεδίο εφαρμογής	Η συνολική διαδικασία από την αρχή μέχρι το τέλος.	Ένα συγκεκριμένο βήμα (το στάδιο της ανάλυσης).
Στόχος	Η εύρεση χρήσιμης γνώσης για τον οργανισμό.	Η αναγνώριση μοτίβων και σχέσεων στα δεδομένα.
Βήματα	Περιλαμβάνει 5 στάδια (καθαρισμό, μετασχηματισμό, κ.λπ.).	Εστιάζει αποκλειστικά στην εφαρμογή αλγορίθμων.
Παρομοίωση	Όλη η επιχείρηση εξόρυξης χρυσού (ψάξιμο, σκάψιμο, καθαρισμός, πώληση).	Το ίδιο το σκάψιμο και το κοσκίνισμα του χώματος για να βρεθεί ο χρυσός.