



Cofinanțat de
Uniunea Europeană



Proiectul Wallachia e-Hub este finanțat prin Digital Europe Programme (EC/101083410) - WeH și prin Programul Creștere Inteligentă, Digitalizare și Instrumente Financiare (POCIDIF/1147/2/1/161799)

GRIGORE ALBEANU,
TRAINER COORDONARE A TRANSFORMĂRII DIGITALE

INOVAȚIA BAZATĂ PE DATE

Știința datelor în acțiune

WEH.SPIRUHARET.RO



EDIH

European
Digital Innovation
Hubs Network

Proiectul

WE HUB

for digital transformation





Datele sunt considerate motorul economiei digitale globale. Inovația bazată pe date este definită ca „o inițiativă strategică a organizațiilor de a utiliza datele și a le analiza pentru a dezvolta perspective bazate pe date care ajută la dezvoltarea de noi produse, îmbunătățirea proceselor, descoperirea de noi piețe și modele de afaceri”. Știința este un studiu sistematic prin observație și experiment. Datele reprezintă fapte și statistici colectate împreună pentru referință sau analiză. Prin urmare, știința datelor constă în explorarea științifică a datelor pentru a extrage sens sau perspectivă și construirea de software pentru a utiliza astfel de informații într-un context al afacerilor/administrației publice.

Prin ȘTIINȚA DATELOR, datele sunt transformate în informații valoroase (vizualizate grafic), produse rezultate din analiza datelor (segmentarea publicului în funcție de anumite reacții), respectiv istorii interesante (reflectate în viitoare strategii de comunicare sau de acțiune). Știința Datelor are la bază tehnici de Ingineria Datelor, utilizează resurse puternice de calcul și metode de analiza datelor (statistică computațională și inferențială, învățare automată, tehnici data mining adaptate la domeniu), tehnici de vizualizare. Toate acestea intervin în cadrul unui proces iterativ specific care constă din: 1) formularea unei ipoteze; 2) colectarea datelor; 3) preprocesarea

datelor (identificare de date incomplete, date multiple, date eronate etc.); 4) Vizualizare și explorare; 5) determinarea modelului; 6) evaluare de metrice și validarea modelului urmată, la nevoie de pașii 2-6, cât timp modelul propus nu îndeplinește cerințele de precizie și potrivire; 7) valorificarea modelului în “producție”.

Pentru evaluarea performanțelor modelelor, utilizarea acurateții modelului general nu este suficientă. Două măsuri sunt cele mai importante: În primul rând, o rată scăzută de descoperire falsă pentru etichetele țintă, pentru a reduce alarmele false. În al doilea rând, scoruri F ridicate pentru toate etichetele țintă, cu excepția celor necunoscute, pentru a asigura o sensibilitate și o reamintire suficiente.

Exemple de utilizare:

- Analiza automată a documentelor cu caracter public și extragerea informațiilor;
- Analiza automată a reacțiilor cetățenilor privind o anumită acțiune/ un anumit subiect;
- Segmentarea populație în funcție de anumite opțiuni/comportamente;
- Personalizarea serviciilor APT în funcție de informația furnizată de aplicațiile de analiza datelor.

