**Curs Data Science Introductiv**

1. **Modul 1 – Introducere in Data Science. Uneltele utilizate in procesul de Data Science**
   1. Ce sunt datele - data vs Big Data. Importanta datelor si impactul cantitatii si calitatii acestora. Motivatia avantului Inteligentei Artificiale, Machine Learning si Data Science din ultimii ani. Exemple.
   2. Modele inferentiale si modele predictive. Statistica vs Data Science. Analiza rezultatelor unor masuratori, sondaje, etc. Erori, matrici de confuzie, variabile corelate. Exemple – fals pozitive vs fals negative in experimente reale (exemple din sanatate si marketing).
   3. Limbaje de programare in Data Science: R, Python, Scala. Comparatii si exemple.
   4. Aplicatii si biblioteci software. Pachete si biblioteci. Medii de executie a experimentelor Data Science.
   5. Medii integrate vs medii de dezvoltare: SPSS/SAS/SAP Hanna vs R Studio, Python, Microsoft Machine Learning, Google Tensor Flow, IBM Watson, etc
   6. Introducerea in modele de regresie. Studiu de caz 1.6: predicitia preturilor in imobiliare. Regresia liniara simpla (uninomiala, univariata)
   7. Introducerea in modele de clasificare. Studiu de caz 1.7: predictia pierderii clientilor. Regresia logistica simpla (uninomiala, univariata)
   8. Introducerea in modele de invatare “nesupravegheata” – clusterizari. Algoritmul kMeans. Studiu de caz 1.8: analiza diverselor specii de plante dintr-o anumita familie si impartirea lor automata in functie de caracteristicile acestora.
   9. Introducere in etapele concrete ale unui experimente de Data Science.
2. **Modul 2 – Introducere in limbajul R**
   1. Introducerea mediului de experimentare R Studio
   2. Stuctura limbajului R
   3. Analiza datelor cu ajutorul obiectului data.frame. Incarcarea acestora din surse diverse.
   4. Proceduri si functii simple in limbajul R
   5. Utilizarea pachetelor de baza R
   6. Vizualizarea datelor din exemplele anterioare (studiile de caz 1.6, 1.7 ,1.8) cu ajutorul R si pachetului ggplot2. Histograme, plotari liniare, plotari cu puncte
   7. Implementarea experimentului de predictie in baza studiului de caz 1.6
   8. Implementarea experimentului de clasificare in limbajul R ( studiul de caz 1.7)
   9. Implementarea experimentului de clusterizare (studiul de caz 1.8)
3. **Modul 3 – Modele reale de analiza predictiva/inferentiala Data Science**
   1. Modele multivariate si multinomiale
   2. Regresie si clasificare multivariata multinomiala
   3. Validarea modelelor – Problema cu doua taisuri a lipsei datelor.
   4. Reluarea experimentelor anterioare si adaugarea validarii si testarii experimentelor (studiile de caz 1.6, 1.7, 1.8)
   5. Modele liniare vs modele non-liniare. Modelarea datelor reale. Exemple concrete.
   6. Alte modele de regresie si clasificare: arbori de decizie, metode bazate pe distante, metode avansate. Introducere in logica si modelele Bayesiene.
   7. Noi metode de regresie si clasificare bazate pe modele ca Decision Tree, k Nearest Neighbors, Support Vector Machines. Extinderea studiilor de caz 1.6, 1.7, 1.8.
   8. Comparatia modelelor.
   9. State-of-the-art in Data Science. Ultimile cercetari si avansuri tehnologie academice si comerciale. Prezentare generala.
4. **Modul 4 – Introducere in procese reale de Data Science**
   1. Culegerea, analiza si curatarea datelor. Exemple si experimente concrete
   2. Preprocesarea datelor. Importanta. Metode. Exemple concrete (referire la experiementele anterioare)
   3. Vizualizarea avansata a datelor
   4. Determinarea importantei variabilelor in modelele inferentiale/predictive. Reducerea si extinderea variabilelor utilizate in desfasurarea unui experiment de Data Science. “Blestemul” dimensionalitatii datelor.
   5. Determinarea modelelor ideale in diverse experimente (extinderea studiilor de caz)
   6. Exemple comparative de experimente utilizand pachete, biblioteci si limbaje diverse
   7. De la experiment Data Science la aplicatie Data Science in mediu de productie. Exemplu concret de operationaliza a experimentelor.