

### **Introduction to Data Science:**

- CRISP-DM methodology, data types and attributes, the similarity of instances (binary, numerical, nominal, ordinal, and mixed instances),
- Clustering and unsupervised learning: definition, hierarchical agglomerative clustering, K-means, DBSCAN, external and internal evaluation of clusters,
- Frequent-pattern mining (itemset, frequent itemset, monotonicity, Apriori, Eclat, FP-Tree, FP- Growth, association rules, quality of a rule),
- Supervised learning: classification, regression, regularization, quality of a supervised model, cross-validation, linear classifiers,
- Recommender systems: types of recommender systems, factorization models, cold-start problem, evaluating recommender systems,
- Data preprocessing: value editing, feature selection, sampling, dimensionality reduction.

### **Machine Learning:**

- Probabilistic methods;
- Bayesian methods;
- Discriminant Functions;
- Parametric models; Probabilistic Graphical models;
- Naive Bayes' Classifier; Cross-validation;
- Hidden Markov models;
- Markov Chain Monte Carlo; Latent Dirichlet Allocation;
- Non-Parametric models;
- Lazy learners (Adaptive Nearest Neighbor Methods, Condensed Nearest Neighbor, Prototype methods, Learning Vector Optimization, Hashing, Locality-sensitive Hashing, k-d trees);
- Decision Trees (Classification trees, Regression trees, Pruning, Multivariate trees, Random Forest);
- Support Vector Machines and Kernel Methods; Regularization; Ensemble techniques, Voting, Stacking, Cascading, Bagging, Boosting;

### **Advanced Deep Network Development (former name: Advanced Machine Learning):**

- Artificial Neural Networks (Artificial neuron, layers, activation functions, learning rate, batch, weights)
- Convolutional Neural Networks (Padding, Stride, Filters, Pooling layer)
- Deep Neural Networks (VGG, AlexNet, R-CNN, U-Net, YOLO, SAM, ResNet, DenseNet, other CNN architectures)
- Embeddings
- Transfer learning
- AutoEncoders (Encoder, Latent Space, Decoder)
- Semantic Segmentation (R-CNN, Fast R-CNN, Faster R-CNN)
- Metrics (IoU, NMS, mAP)
- Recurrent Neural Networks (LSTM, GRU, Seq2Seq, Bidirectional, Attention mechanism)
- Transformers

**Stream mining:**

- Part I: Stream mining 101
  - Querying streams
  - The use of hashing in stream mining
  - Repeated search
  - Sampling
  - Filtering
  - Hash functions
  - The Bloom filter
  - Frequency of elements
  - Frequency moments
  - Clustering types in data streams
  - Hierarchical clustering
  - Naïve bayes in data streams
  - Decisions trees in data stream classification
  - Hoeffding tree
- Part II: Stream mining 202
  - Sketch creation process
  - Sketches in counting elements and frequency problems
  - Sketch mergeability
  - Frequent patterns in data streams
  - Sub- & super-patterns + closed & maximal patterns
  - Coresets
  - Sequential pattern mining
  - Time series categorization
  - Trends, seasonality and similarity
  - Clustering, classification and anomaly detection in data streams
  - The change management process
  - Rate of change
  - Blind vs adaptive change management
- Part III: Streaming systems
  - Non-grouping and grouping operations
  - The Lambda architecture
  - Event and processing time windowing
  - Watermarks, allowed lateness
  - Windowing & sessions
  - Triggers
  - At-most-once, at-least-once & exactly-once guarantees
  - Checkpoints and message logging in streaming systems
  - Stream processing platforms (Kafka, Flink, etc) and consistency

**Open-source technologies:**

- Data storage solutions: MongoDB, Logstash, Cassandra, HDFS, HBase
- Data analysis solutions: MapReduce, Spark, Elasticsearch
- Streaming systems: Kafka, Storm, Spark Streaming, Flink
- Data visualization: viz types, Tableau, Datawrapper, Plotly
- Other topics: Docker