SYLLABUS FOR BOARD OF STUDIES ADOPTION — generated from the live catalogue record

# Machine Learning with Python and scikit-learn

S-03A · 5 days (configurable 1–5 days) · Informatics · AI · Years: 3rd, PG

## CONTENTS

1. Overview 2
2. Hours and credit position 2
3. Course outcomes 2
4. Course content — modules 2
5. Assessment scheme (100 marks) 3
6. CO–PO mapping 3
7. Every participant receives 3
8. Fee 3
9. Designed by 3

Catalogue S-03A · version 2 · generated 2026-10-11

## OVERVIEW

End-to-end: data preparation, model selection, evaluation and a deployed prediction service.

Machine learning is the fastest-growing skill demand across sectors; an end-to-end project from data to a deployed model gives students a portfolio piece and genuine competence.

## HOURS AND CREDIT POSITION

Contact hours: 40 Guided project: 20 Self-study: 0 Notional total: 60

**Credit claim: 2**
60 notional hours ÷ 30 hours per credit (NSQF track, UGC Micro-credentials Guidelines §5.3.3) = 2 credits.

## COURSE OUTCOMES

CO1. Prepare a raw dataset for modelling through cleaning, encoding and scaling (K3)
CO2. Build supervised regression and classification models with scikit-learn (K3)
CO3. Evaluate a model using cross-validation and appropriate metrics (K4)
CO4. Tune hyperparameters to reduce overfitting (K4)
CO5. Deploy a trained model as a prediction service (K3)

## COURSE CONTENT — MODULES

Hours shown are the recommended format; module time scales proportionally when the engagement runs 1–5 days.

**Module 1 — ML foundations and workflow (5h recommended)**
Theory: Problem framing, the ML pipeline, train/test discipline
Lab: Frame a problem and split a dataset correctly
Output: Problem-framing note

**Module 2 — Data preparation (5h recommended)**
Theory: Cleaning, encoding, scaling, feature engineering
Lab: Prepare a raw dataset for modelling
Output: Prepared dataset + notebook

**Module 3 — Regression (5h recommended)**
Theory: Linear and regularised regression, evaluation metrics
Lab: Build and evaluate a regression model
Output: Regression notebook

**Module 4 — Classification (5h recommended)**
Theory: Logistic regression, decision trees, k-NN, support vector machines
Lab: Build and compare classifiers
Output: Classification notebook

**Module 5 — Evaluation and tuning (5h recommended)**
Theory: Cross-validation, hyperparameter search, over- and underfitting
Lab: Tune a model with cross-validation
Output: Tuning report

**Module 6 — Ensemble methods (5h recommended)**
Theory: Random forests, gradient boosting
Lab: Improve results with ensembles
Output: Ensemble notebook

**Module 7 — Unsupervised learning (5h recommended)**
Theory: Clustering, dimensionality reduction
Lab: Cluster a dataset and interpret the groups
Output: Clustering notebook

**Module 8 — Deployment capstone (5h recommended)**
Theory: Packaging a model, serving predictions
Lab: Deploy the best model as a service
Output: Deployed prediction service

## ASSESSMENT SCHEME (100 MARKS)

Continuous lab exercises — 40 marks
Capstone project & demonstration — 40 marks
Quiz & viva voce — 20 marks

## CO–PO MAPPING

CO1 !’ PO3 (strength 2)
Data preparation is solution development.
CO2 !’ PO3 (strength 2)
Model building is solution development.
CO3 !’ PO4 (strength 2)
Model evaluation is experimental investigation.
CO4 !’ PO4 (strength 2)
Tuning is investigation of model behaviour.
CO5 !’ PO5 (strength 2)
Deployment is modern-tool usage.

## EVERY PARTICIPANT RECEIVES

- Certificate of completion
- Course material
- LMS access
- Interview question bank
- Mock interview & viva practice
- Optional nasscom NSQF assessment (priced per candidate on your quotation)

## FEE

¹1,500 per student, incl. GST
Excludes the optional nasscom assessment where offered. Final pricing on your college's quotation.

## DESIGNED BY

Venkat Kurela — Founder & Director, Kurela Cognisive Private Limited