Workshop on Machine Learning in Survival Analysis with Hands-on R (WMLSAR)

Syllabus to be covered (Tentative)

Prerequisites:

- Probability & Mathematical Statistics
- Linear Models and Regression (preferred but not necessary)
- Basic Survival Analysis (preferred but not necessary)
- Introductory Machine Learning (preferred but not necessary)

Course Description

This course integrates **classical survival analysis** with **modern machine learning techniques** for time-to-event data. The workshop will begin with the foundations of survival analysis, followed by more advanced topics including **regularized Cox models, survival trees, random survival forests, gradient boosting, deep learning models, competing risks, multi-state models, and causal survival ML**. Applications span **healthcare, reliability engineering, finance, and social sciences**.

Course Objectives

By the end of the course, students will be able to:

- Understand censoring and truncation in complex ML settings
- Apply ML algorithms adapted for survival data
- Compare classical and ML-based survival models
- Implement survival ML models in **R**
- Interpret survival ML models responsibly
- Conduct model validation using survival-specific metrics

Module-wise Syllabus

Module 1: Foundations of Survival Analysis

- Survival time, censoring, truncation
- Survival function, hazard, cumulative hazard
- Kaplan–Meier estimator
- Nelson–Aalen estimator
- Log-rank test
- Cox Proportional Hazards Model
- Partial likelihood
- Time-dependent covariates

Module 2: Limitations of Classical Survival Models

- Proportional hazards assumption
- Non-linearity and interactions
- High-dimensional covariates
- Model misspecification
- Motivation for ML approaches

Module 3: Regularized Survival Models

- Penalized Cox models
  - LASSO
  - Ridge
  - Elastic Net
- Variable selection in survival data
- High-dimensional survival analysis
- Cross-validation for censored data

Module 4: Survival Trees and Ensembles

- Survival decision trees
- Splitting criteria (log-rank, likelihood-based)
- Random Survival Forests (RSF)
- Variable importance measures
- Conditional inference trees

Module 5: Boosting Methods for Survival Data

- Gradient boosting for survival analysis
- CoxBoost
- Accelerated Failure Time (AFT) boosting
- XGBoost for survival data
- LightGBM survival extensions
