

2025年第四届“创新杯”（原钉钉杯）大学生大数据挑战赛题目

初赛A：智慧工厂工业设备传感器数据分析

一、赛题背景:

智慧工厂是通过数字化、智能化和自动化技术深度融合，重塑制造业的生产模式、管理流程和价值链，推动工业升级和可持续发展。通过工业机器人、AGV（自动导引车）等设备实现重复性工作的自动化，减少人工干预，提高生产速度和一致性。利用大数据和AI算法实时监控设备状态、物料消耗和生产进度，动态调整生产计划，减少资源浪费（如能源、原材料）。自动化替代人工完成危险或高强度任务（如焊接、搬运），同时减少对低技能劳动力的依赖。

参赛者将获得一份工厂传感器模拟器数据集，这是一个大规模的合成数据集，专为工业5.0背景下的预测性维护、异常检测和工业机器学习应用而设计。该数据集包含部署在未来智能工厂环境中的50万台模拟机器的传感器读数、操作指标和维护记录。

该数据集包含30多个真实场景的机器设备类型（数控铣床，熔炉，机械臂，激光切割机），核心传感器数据：温度，振动，声音，功率，油/冷却液液位；维护历史和人工智能监督领域；机器特定的功能，如测量激光强度，液压压力等。

二、样本解读:

Column Name名称	Description特征解释
Machine_ID	机器编号
Machine_Type	机器类型
Installation_Year	安装年份
Operational_Hours	运行小时数
Temperature_C	温度(°C)
Vibration_mms	振动(毫米/秒)
Sound_dB	声音(分贝)
Oil_Level_pct	油位(百分比)
Coolant_Level_pct	冷却液位(百分比)
Power_Consumption_kW	功耗(千瓦)
Last_Maintenance_Days_Ago	距上次维护天数
Maintenance_History_Count	维护历史次数
Failure_History_Count	故障历史次数

AI_Supervision	人工智能监控
Error_Codes_Last_30_Days	过去30天错误编码
Remaining_Useful_Life_days	剩余使用寿命(天)
Failure_Within_7_Days	7天内故障预测
Laser_Intensity	激光强度
Hydraulic_Pressure_bar	液压压力
Coolant_Flow_L_min	冷却液流量(升/分钟)
Heat_Index	热指数
AI_Override_Events	人工智能覆盖率

三. 解决问题：

初赛任务A：机床设备故障预测回归分析问题

使用机器编号、机器类型、运行小时数、温度、振动、声音、油位、冷却液位、功耗、距上次维护天数、维护历史次数、故障历史次数、人工智能监控、过去30天错误编码等特征，构建回归模型，预测机床设备在7天内是否会发生故障。要求输出模型准确率、召回率、F1值，并分析前5个最重要的特征。

初赛任务B：剩余使用寿命预测回归分析问题

在不使用目标标签（7天内故障预测）的情况下，基于剩余使用寿命（天）、运行小时数、温度、振动、油位、冷却液位、维护历史次数、故障历史次数等特征，构建回归模型，预测机床设备的剩余使用寿命（目标为连续值）。要求输出模型的均方误差（MSE）和决定系数（R²）并分析特征重要性对剩余寿命的影响。

题目声明：本赛事所有赛题仅授权 2025年第四届“创新杯”（原钉钉杯）大学生大数据挑战赛参赛队伍使用，任何组织及个人未经组委会书面授权，严禁用于校内竞赛，篡改、复制等侵权行为。