**重庆邮电大学本科毕业设计（论文） 任务书**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| 题 目 | 基于ADASYN和机器学习模型的水质检测系统研发 | | | | |
| 学生姓名 | 刘畅 | | | 学 号 | 2020215139 |
| 指导教师 | 许汀汀 | | | 所在单位 | 软件工程学院重邮-惠普软件学院 |
| 题目类型 | 应用型√ 研究型 综合型 其它 | | | | |
| 是否需要在实验、实习、工程实践和社会调查等社会实践中完成（√是 否 ） | | | | | |
| 一、研究目标 | | | | | |
| 1. 开发ADASYN算法应用于水质检测： 实现自适应合成采样算法(ADASYN)在水质检测中的应用，通过该算法生成综合阈值以上的粪便指示细菌(FIB)实例。该目标的达成将有助于平衡数据集，提高水质检测系统对异常情况的适应能力，从而增加系统的鲁棒性。 2. 集成水质检测的机器学习模型： 设计增强决策树、支持向量机(SVM)和多层感知器人工神经网络(MLP)等机器学习模型的集成框架。通过选择和优化合适的模型，使其能够有效地对FIB实例进行分类和预测，提高水质检测系统的准确性。 3. 实现综合水质检测系统： 将ADASYN算法与机器学习模型相结合，建立一个综合水质检测系统。该系统能够自动地从水样本中提取特征、生成综合阈值以上的FIB实例，并利用不同的机器学习模型进行水质的评估和预测。 | | | | | |
| 二、主要研究内容和方法 | | | | | |
| * 研究内容： * **解决数据的不平衡问题：** 针对水质监测数据中的不平衡性，即绝大多数样本属于正常水质，而异常样本（FIB实例）相对较少的问题，通过引入ADASYN算法来生成合成的异常样本，以平衡数据集。 * **机器学习模型集成及优化：** 研究如何集成和选择增强决策树、支持向量机(SVM)和多层感知器人工神经网络(MLP)等机器学习模型，以适应水质监测任务。考虑到每个模型的特性，对其参数进行调优，使其在水质检测中能够达到最佳性能。 * **系统集成与研发：** 将ADASYN异常样本算法模型与机器学习模型相结合，设计并构建一个综合水质检测系统。通过优化系统参数和整合各个模块，实现整个系统的协同工作，提高水质检测的综合性能。 * 研究方法： * **ADASYN算法：** 对水质监测数据进行分析，确定异常样本的分布情况，然后采用ADASYN算法生成合成的异常样本，以解决数据不平衡问题。关键在于合理设置ADASYN算法的参数，确保生成的样本既能够平衡数据，又不引入过多的噪音。 * **机器学习模型：** 通过实验比较增强决策树、支持向量机(SVM)和多层感知器人工神经网络(MLP)的性能，选取在水质监测中表现最优的模型。然后，通过交叉验证等方法对选定的模型进行参数调优，提高其在水质检测任务中的准确性。 | | | | | |
| 三、主要考核要求或指标 | | | | | |
| **主要考核指标：系统平台+论文**  **平台功能要求:**   1. **数据平衡和合成效果：**  * **功能要求：** ADASYN算法能够有效平衡水质监测数据集，生成具有代表性的粪便指示细菌(FIB)实例。 * **技术指标：** 数据集中正常水质和异常水质样本的比例应接近理想的平衡状态，生成的合成异常样本应具有高度可信度。  1. **机器学习模型性能：**  * **功能要求：** 机器学习模型（增强决策树、支持向量机、多层感知器人工神经网络）在水质检测任务中能够高效、准确地分类和预测。 * **技术指标：** 模型在测试数据集上的准确度、精确度、召回率等性能指标应满足预定的要求。  1. **系统综合性能：**  * **功能要求：** 水质检测系统能够集成ADASYN生成的样本和机器学习模型，实现对水质异常的综合监测和预测。 * **技术指标：** 整个系统的预测准确度、实时性、鲁棒性等性能指标应达到设计目标。  1. **论文质量和创新点：**  * **功能要求：** 撰写规范、清晰、有逻辑的论文，准确反映研究过程和结果，具有一定的创新性。 * **技术指标：** 论文应包括对ADASYN算法的详细解释、机器学习模型的选择和优化过程、系统设计和实验结果等，论文结构合理，创新点明确。 | | | | | |
| 四、主要参考文献 | | | | | |
| 1. Xu T, Coco G, Neale M. A predictive model of recreational water quality based on adaptive synthetic sampling algorithms and machine learning[J]. Water research, 2020, 177: 115788. 2. Zhu M, Wang J, Yang X, et al. A review of the application of machine learning in water quality evaluation[J]. Eco-Environment & Health, 2022. 3. Ahmed A N, Othman F B, Afan H A, et al. Machine learning methods for better water quality prediction[J]. Journal of Hydrology, 2019, 578: 124084. 4. Devane M L, Moriarty E, Weaver L, et al. Fecal indicator bacteria from environmental sources; strategies for identification to improve water quality monitoring[J]. Water Research, 2020, 185: 116204. 5. He H, Bai Y, Garcia E A, et al. ADASYN: Adaptive synthetic sampling approach for imbalanced learning[C]//2008 IEEE international joint conference on neural networks (IEEE world congress on computational intelligence). Ieee, 2008: 1322-1328. | | | | | |
| 指导教师签字：许汀汀 | | | 2023 年01 月28日 | | |
| 专业负责人意见：  □同意立题  □不同意立题   |  |  |  |  |  |  | | --- | --- | --- | --- | --- | --- | | 负责人签字： | |  | | | | |  | 年 |  | 月 |  | 日 | | | 学院意见：  □同意立题  □不同意立题   |  |  |  |  |  |  | | --- | --- | --- | --- | --- | --- | | 负责人签章： | |  | | | | |  | 年 |  | 月 |  | 日 | | | | |

备注：1.此任务书由指导教师填写，并于毕业设计（论文）选题结束后尽快下达给学生。

2.任务书一经审定，指导教师和学生不得随意更改，如因特殊情况确需变更，应在完成开题报告之前，填写《重庆邮电大学毕业设计（论文）更改题目审批表》，报专业负责人审核、学院复核批准后执行，并报教务处备案。