

**毕业设计(论文)教学档案**

|  |  |
| --- | --- |
| 题 目： | 基于ADASYN和机器学习模型的水质检测系统研发 |
| 姓 名： | 刘畅 |
| 学 院： | 国际学院 |
| 专 业： | 软件工程 |
| 班 级： | 34082003 |
| 学 号： | 2020215139 |
| 指导老师： | 许汀汀 |

**重庆邮电大学教务处制**

**二〇二四年六月**

目录

1. 重庆邮电大学本科毕业设计（论文）任务书；
2. 重庆邮电大学本科毕业设计（论文）开题报告；
3. 重庆邮电大学本科毕业设计（论文）工作周报；
4. 重庆邮电大学本科毕业设计（论文）中期检查表；
5. 重庆邮电大学本科毕业设计（论文）指导教师评语表；
6. 重庆邮电大学本科毕业设计（论文）评阅意见表；
7. 重庆邮电大学本科毕业设计（论文）答辩记录表；
8. 重庆邮电大学本科毕业设计（论文）答辩委员会意见及评分表；

**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. | | | | |
| 图片包含 文本描述已自动生成指导教师签字： 2023 年12 月18日 | | | | |
| 专业负责人意见：  □同意立题  □不同意立题   |  |  |  |  |  |  | | --- | --- | --- | --- | --- | --- | | 负责人签字： | |  | | | | | 2023 | 年 | 12 | 月 | 20 | 日 | | | 学院意见：  □同意立题  □不同意立题   |  |  |  |  |  |  | | --- | --- | --- | --- | --- | --- | | 负责人签章： | |  | | | | | 2023 | 年 | 12 | 月 | 22 | 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。

**重庆邮电大学本科毕业设计（论文）开题报告**

|  |  |  |  |
| --- | --- | --- | --- |
| 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 |
| 一、选题背景（综述本课题研究现状、选题目的及意义） | | | |
| 随着工业化和城市化进程的加快，水质污染日益严重，给人类健康和环境造成了严重的威胁。因此，开展水质检测工作具有重要的现实意义。近年来，利用人工智能和机器学习技术解决水质检测问题成为了研究的热点之一。传统的水质监测方法往往需要大量的人力和物力投入，并且存在着检测周期长、成本高等问题。而利用机器学习模型和算法进行水质监测，则能够实现自动化、高效率的监测，大大提高了水质监测的效率和准确性。  当前，水质检测领域的研究主要集中在解决数据不平衡问题、选择和优化机器学习模型以及构建综合水质检测系统等方面。在数据不平衡问题上，研究者们提出了各种方法，如过采样、欠采样等，但在异常样本数据稀缺的情况下仍有局限性。在机器学习模型的选择和优化方面，尽管已尝试多种模型，但如何选择合适的模型并进行优化仍存在争论。  未来，水质检测技术将朝着智能化、自动化、高效化的方向发展。随着人工智能和大数据技术的成熟和应用，我们有望在数据不平衡问题、机器学习模型优化以及系统构建等方面取得更多创新和突破。 | | | |
| 二、研究目标和内容 | | | |
| 2.1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。   2.2主要研究内容   * 1. **前端搭建：** * 使用Vue.js搭建前端应用，通过Vue CLI工具创建项目。 * 结合Element UI，利用其提供的丰富组件快速搭建界面。 * 设计界面包括数据展示、算法参数设置等功能，确保用户友好的交互体验。   1. **后端搭建：** * 使用Python语言进行后端开发，通过Flask框架建立RESTful API。 * 设计API接口，包括数据获取、算法调用和结果返回等功能。 * 在后端使用Python实现ADASYN算法，用于合成数据集。 * 确保ADASYN算法的高效性和准确性，以提升数据集的平衡性。 * 利用Python的机器学习库（如Scikit-learn、TensorFlow等）实现增强决策树、支持向量机和多层感知器等模型。 * 通过交叉验证和调参等方法优化模型，以提高水质检测系统的准确性和鲁棒性。   1. **前后端交互：**      + - 设计统一的API接口，定义请求和响应的数据格式。        - 使用Flask提供的路由功能，将前端发送的请求映射到相应的处理函数上。        - 在Vue.js中使用Axios等HTTP客户端库发送请求到后端API。        - 接收后端返回的数据，并进行相应的处理和展示。        - 使用JSON格式进行前后端数据的传输，确保数据的格式统一和易于处理。 | | | |
| 三、研究方案 | | | |
| 3.1研究方法  1. 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。  2. 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。  3.2实施步骤  1. 数据收集与准备：收集水质监测数据集，包括正常水质和异常样本（FIB实例）。对数据进行清洗、预处理和特征提取，为后续的模型训练做准备。  2. ADASYN算法实现与应用：基于收集到的数据，实现ADASYN算法，并将其应用于不平衡数据集，生成合成的异常样本。调整算法参数以确保生成的样本能够有效地平衡数据集。  3. 机器学习模型选择与优化：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，并针对每个模型进行参数调优和性能评估。通过交叉验证等技术，选择最优的模型组合。  4. 模型集成与系统构建：将经过优化的机器学习模型与ADASYN算法相结合，构建一个综合的水质检测系统。设计系统的整体架构，并实现各个模块之间的协同工作，确保系统能够有效地进行水质检测和评估。  5. 系统测试与性能评估：对构建的水质检测系统进行全面的测试和评估。使用真实的水质监测数据对系统进行验证，并对系统的准确性、鲁棒性和性能进行评估和分析。  3.3拟解决的主要问题及措施  1. 数据不平衡问题：通过引入ADASYN算法生成合成的异常样本，以平衡数据集，解决水质监测数据中正常样本和异常样本不平衡的问题。  2. 模型选择与优化：针对水质监测任务，选择合适的机器学习模型，并对其参数进行调优，以提高模型的分类和预测性能。  3. 系统集成与优化：将ADASYN算法与机器学习模型集成到一个统一的水质检测系统中，设计合适的系统架构，并优化各个模块之间的协同工作，以提高系统的综合性能和效率。 | | | |
| 三、进度计划（按月编制） | | | |
| |  |  |  | | --- | --- | --- | | 时间 | 主要工作 | 预期阶段成果 | | 2024年1月 | 项目启动：明确研究方向，创新点及问题定义。  文献综述：   1. Xu, Tingting & Coco, Giovanni & Neale, Martin. (2020). A predictive model of recreational water quality based on adaptive synthetic sampling algorithms and machine learning. Water Research. 177.115788.10.1016/j.watres.2020.115788. 2. Haibo He, Yang Bai, E. A. Garcia and Shutao Li, ADASYN: Adaptive synthetic sampling approach for imbalanced learning, 2008 IEEE International Joint Conference on Neural Networks 2008, pp.1322-1328,10.1109/IJCNN.2008.4633969. 3. Searcy RT, Phaneuf JR, Boehm AB. High-frequency fecal indicator bacteria (FIB) observations to assess water quality drivers at an enclosed beach. PLoS One. 2023 Jun 2;18(6):e0286029. doi: 10.1371/journal.pone.0286029. PMID: 37267238; PMCID: PMC10237476. | 确定研究方向和问题定义，完成文献综述 | | 2024年2月 | 数据收集与准备，包括水质监测数据的收集和清洗预处理。 | 收集整理水质监测数据，完成数据清洗和预处理 | | 2024年3月 | 实现ADASYN算法并应用于数据集，进行数据平衡处理。 | 完成ADASYN算法的实现和应用，生成平衡的数据集 | | 2024年4月 | 选择并优化机器学习模型，包括增强决策树、SVM和MLP等。 | 完成机器学习模型的选择和参数调优 | | 2024年5月 | 构建综合水质检测系统，集成ADASYN算法和机器学习模型。 | 完成水质检测系统的搭建和集成，实现整体功能 | | | | |
| 四、指导教师意见 | | | |
| □同意开题  □不同意开题   |  |  |  |  |  |  | | --- | --- | --- | --- | --- | --- | | 指导教师签字： | |  | | | | | 2024 | 年 | 1 | 月 | 3 | 日 | | | | |

备注：此报告应根据下达的毕业设计(论文)任务书，在指导教师的指导下由学生独立撰写，并于任务书下达后两周内完成。

**重庆邮电大学本科毕业设计（论文）工作周报**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 基  本  信  息 | 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 |
| 工作进展 | 导师指导学生对水质检测领域的相关文献进行了全面的综述，包括数据不平衡处理方法、机器学习模型在水质检测中的应用以及综合水质检测系统的构建等方面的研究进展。学生列举了相关的经典文献和研究成果，对各种方法的优缺点进行了分析比较，并指出了目前研究中存在的问题和挑战。在综述的基础上，学生和导师一起讨论了毕业设计论文的具体方向和内容。通过深入交流，确定了以过采样算法和机器学习模型在水质检测中的应用为主要研究内容，同时结合实际情况构建综合水质检测系统的方向。 本次指导主要是确定选题。该生一直跟随指导老师进行科研活动，所以很快就确定了毕业设计题目。之前的研究偏理论方法，所以在本次的知道中帮助学生将理论方法类研究转换成为软件工程类的实现，确定了毕业设计工作的方向。给出了文献供该学生进行扩展阅读。 | | | |
| 工作进展 | 导师指导学生学习实现ADASYN算法，并将其应用于水质检测数据集，生成合成的异常样本，以解决数据不平衡问题。在这个过程中，导师提供了相关文献资料和算法实现的指导，帮助学生理解ADASYN算法的原理和应用场景，并指导学生将算法应用于具体的水质检测数据集中，生成合成的异常样本以实现数据集的平衡化处理。此外，导师还指导学生学习进一步的研究计划，包括优化ADASYN算法的实现，探索其他数据不平衡处理方法，以及选择合适的机器学习模型进行后续实验。例如，导师鼓励学生尝试使用KNN、GDBT等机器学习模型，以确定最适合水质检测的模型，并进行性能评估和比较。通过导师的指导和支持，学生得以系统地学习和应用相关算法和方法，为毕业设计的顺利进行提供了重要的支持和指导。 本次指导关注算法模型。学生虽然接触过深度学习算法，但是对数据生成算法还不属性。在指导过程当中主要给学生讲解了两个经典的数据生成算法的思路和内涵，该生的学习能力很强，通过本次指导基本就掌握了该类算法的思路，为毕业设计打好了基础。希望下来可以尽快编写代码，落实算法实现。 | | | |
| 工作进展 | 导师指导学生选择合适的机器学习模型，包括增强决策树、支持向量机和多层感知器等；此外，导师针对选定的机器学习模型，指导学生进行了模型优化的工作。包括调整超参数、优化模型结构、特征选择等方面的工作，以提高模型的性能和效率，以提高水质检测系统的准确性和效率，同时帮助学生制定了进一步优化和应用机器学习模型的计划，包括进一步细化模型参数、探索其他机器学习模型、加入特征工程等方面的工作。 因为本毕业设计涉及多个机器学习模型，所以借着这个机会给组内所有的学生普及了一下四个经典的机器学习模型。学生们对此内容都很感兴趣，大家认真的做了笔记和记录。 | | | |
| 工作进展 | 导师指导学生根据设计稿，完成了水质检测系统的各个页面的开发工作。包括数据上传、数据可视化展示、模型训练和预测等功能的实现，确保了系统的完整性和功能性。建议学生选择了Vue3作为前端框架，并结合Element UI等组件库进行开发。通过Vue3的响应式数据绑定和组件化开发特性，实现了界面的动态更新和交互功能。在后期测试工作中， 帮助学生进行了系统的测试工作，发现并修复了一些界面bug和功能问题，并积极收集用户反馈，针对性地进行了界面改进和功能优化。 主要指导学生完善了毕业设计开发的系统。学生有很强的动手能力，基本按照要求完成了系统实现。但是整个系统的操作性还比较差，界面不美观，所以又指导学生对系统的一些细节进行的调整和优化。 | | | |
| 工作进展 | 导师指导学生采用Flask框架进行后端模块的设计和开发。通过Flask提供的路由、视图函数等功能，实现了与前端Vue框架的交互，并确保了系统的稳定性和可扩展性。帮助学生设计了模型部署的接口和逻辑，通过接口调用，使用训练后的最优模型对新数据的预测分析，并将结果返回给前端界面。并同学生进行了后端模块的测试工作，发现并修复了一些bug和功能问题。并根据系统需求和性能要求，优化了代码结构和算法实现，提高了系统的效率和稳定性。 主要指导学生进行了算法模型与系统的接口融合，进行了测试，学生很好的完成了布置的任务，按照毕业设计计划进行工作。在测试的时候发现了一些数据问题，进行了修复。同时发现学生开发的系统界面冗余很多，没有体现主要功能，也安排学生进行了修改。 | | | |
| 工作进展 | 导师对论文的各个章节进行了全面审核，确保研究背景、文献综述、研究方法、数据分析、结论等部分齐全，并且相互呼应，形成完整的论文结构。同时，详细检查了论文中的理论引用，确保引用文献的权威性和相关性，并指导学生如何更好地运用理论支持自己的研究观点。此外，导师指出了论文中研究方法部分的不足，特别是数据收集和分析方法的详细描述，并提供了改进建议，确保方法部分的描述准确、详细。最后，针对论文中的语言表达和格式问题，导师提供了详细的指导，包括学术语言的使用、引用格式的规范、图表的制作等，提升论文的整体质量和学术规范性。 学生基本完成了论文初稿，所以本次指导主要对论文进行详细的指导。首先对整个论文的架构进行了检查，发现系统需求和设计方面内容还较少。同时对格式规范进行了检查，基本符合要求。对于论文出现的问题也要求学生马上进行整改完善。 | | | |
| 工作进展 | 导师强调了答辩时间的合理分配，并指导如何在规定时间内高效、简洁地表达研究成果，提供了关于应对可能提问的策略，如何突出论文亮点及研究创新点，确保答辩时能自信从容地回答评委问题。进行了模拟答辩，通过角色扮演模拟评委提问，帮助学生熟悉答辩流程和提高临场应变能力，同时对答辩过程中发现的论文内容逻辑问题进行详细讲解，强调论点之间的联系和过渡，确保论述严谨有序。本次指导主要和小组学生进行了预答辩。重点检查了答辩PPT和论文修改情况。该生的论文基本符合要求，但是PPT中没有需求分析，详细设计的内容也偏少，要求该学生添加上述内容，并模拟了一些问题，学生能够很好的回答，说明其对整个毕业设计任务完成质量高。同时对论文存在的一些细节问题也进行了指导修正。 | | | |
| 说明：此表是由毕业设计系统直接导出。来源于由学生和指导老师共同填写的《重庆邮电大学本科毕业设计（论文）工作周报》. | | | | |

**重庆邮电大学本科毕业设计（论文）中期检查表**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 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 |
| 学生填写 | 目前已完成任务概述：  1主要内容: (毕业设计（论文）进展情况，字数一般不少于500字)   1. 系统架构设计：完成了前端和后端的架构设计，前端采用Vue3和element UI进行开发，在前端页面中展示了水质的采样地点，衡量水质的指标，同时实现了数据上传、算法选择、模型训练等功能，并提供了友好的用户界面。后端采用Django框架实现，对前端接口进行分析，返回模型处理结果，实现了前后端之间的数据交互和通信，确保了系统的正常运行和数据的准确传输。 2. 数据集获取与处理：成功获取了来自Kaggle平台的水质数据集，并进行了数据预处理和特征工程，选取了PH值，导电率，温度等作为特征值，为后续的机器学习模型训练做好了准备，并在前端界面显示数据的具体内容。 3. 过采样算法应用：在前端界面引入了ADASYN和SMOTE算法，通过增加少数类样本的数量，解决了数据不平衡的问题，使得水质数据集的不同类别的水质数据量相同，为后续模型训练，提供了可靠的数据集，提高了模型的性能。 4. 机器学习模型训练：利用KNN、GDBT、SVM和ANN等机器学习模型对过采样后的数据进行了训练，并初步评估了模型的性能，将模型的准确率在前端界面加以显示，用户可以根据需求选择合适的机器学习模型进行分析及预测新数据的水质情况。 5. 结果可视化展示：通过图表和可视化工具，展示了模型训练结果、数据分布情况以及算法性能指标，例如准确率，使用户可以直观地了解系统的运行情况。   2.尚存在的问题及采取的措施：   1. 特征选择不充分：目前只选取了少数特征进行分析，未能充分利用数据集中的所有信息。解决措施是进一步探索数据集，选择更具代表性的特征进行分析。 2. 高维数据处理困难：对于高维数据，SVM模型训练过程中的计算时间较长，影响了系统的实时性。解决措施是优化算法实现，引入调优过程，针对不同的机器学习模型，进行了超参数的调优和性能优化，以提高模型的准确性和泛化能力。 3. 模型评估不完善：目前系统对于模型的评估仅限于准确率等简单指标，缺乏更深层次的评估。解决措施是引入更多的评估指标，如召回率、精确率和F1值等，来全面评估模型的性能。 | | | |
| 指  导  教  师  填  写 | 学生调研及查阅文献情况 | | □优□良□合格□不合格 | |
| 毕业设计（论文）原计划是否调整 | | □是□否 | |
| 学生是否按计划执行工作进度 | | □是□否 | |
| 学生是否能独立完成工作任务 | | □是□否 | |
| 学生的英文翻译是否按进度进行 | | □是□否 | |
| 学生的工作态度 | | □认真□一般□较差 | |
| 对调整计划的意见（若计划有调整，说明原因） | | | |
| 指导教师意见：该生学习态度积极认真，能够很好的完成既定的毕业设计任务，同时能够对现有任务进行扩展，提出表扬，有较好的动手能力，基本上完成了项目75%以上的内容，同意通过中期检查。现阶段的系统还不完善，特别是在数据特征的选择和整个系统的展示上还有待提高。建议学生对数据集数据的特征展开深入的研究，同时对系统展示，特别是进入系统后的首页展示进行进一步的优化调整。   |  |  |  |  |  |  | | --- | --- | --- | --- | --- | --- | | 指导教师签字： | |  | | | | | 2024 | 年 | 4 | 月 | 1 | 日 | | | | |
| 检查小组审核 | 审核意见：   |  |  |  |  |  |  | | --- | --- | --- | --- | --- | --- | | 负责人签字： | |  | | | | | 2024 | 年 | 4 | 月 | 5 | 日 | | | | |

**重庆邮电大学本科毕业设计（论文）指导教师评语表**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| 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 | |
| 评分项目 | | | | | 满分分值 | 评定得分 | |
| 1 | 调查研究 | | | | 10 | 9 | |
| 2 | 方案设计 | | | | 20 | 16 | |
| 3 | 任务完成情况及论文写作质量 | | | | 50 | 44 | |
| 4 | 工作态度和表现 | | | | 10 | 9 | |
| 5 | 计算机、外语运用能力 | | | | 10 | 8 | |
| 指导教师评定成绩（百分制） | | | | | | 86 | | |
| 指导教师评语 | | 从工作态度和表现、所做的工作及成果情况、论文写作质量等方面进行评价，最后给出结论性评价和意见（达到/基本达到/未达到毕业论文要求，同意/不同意提交毕业论文答辩）  本毕业论文格式规范，该通过详细的研究和实现，成功构建了一个基于ADASYN和机器学习技术的水质检测系统，在水质监测领域具有较高的应用价值。论文结构合理，内容详实，为相关领域提供了一个有效的参考和借鉴，满足软件工程本科生专业毕业要求，同意该生参与毕业答辩。文章详细介绍了ADASYN和SMOTE等自适应过采样算法，以解决数据不平衡问题，同时使用KNN、GDBT、SVM和ANN等机器学习模型对平衡后的数据进行训练和预测，并在此基础上形成了一个水质检测系统。并利用公共数据集对此系统进行了应用验证。接下来可以考虑的是论文中先使用的数据集规模较小，可能影响模型的泛化性能和实际应用效果，需要增加数据集；系统的用户界面设计和交互体验可以进一步优化，如增加更多的可视化工具和友好的错误提示。   |  |  |  |  |  |  | | --- | --- | --- | --- | --- | --- | | 指导教师签字： | | 图片包含 文本描述已自动生成 | | | | | 2024 | 年 | 5 | 月 | 22 | 日 | | | | | | | |
|  | | | | | | |

**重庆邮电大学本科毕业设计（论文）评阅意见表**

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 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 | | |
| 评分项目 | | | | | | 满分分值 | 评定得分 | |
| 1 | 选题符合专业培养目标；体现综合训练基本要求；理论意义和实用价值 | | | | | 15 | 13 | |
| 2 | 查阅文献资料能力；综合运用知识能力；研究方案的设计能力；研究方法和手段的运用能力；外文运用能力 | | | | | 25 | 20 | |
| 3 | 题目难易度；工作量； | | | | | 25 | 18 | |
| 4 | 写作水平、写作规范 | | | | | 15 | 9 | |
| 5 | 研究成果的创新性 | | | | | 10 | 7 | |
| 6 | 成果的理论或实用价值 | | | | | 10 | 8 | |
| 评阅人评定成绩（百分制） | | | | | | | 75 | |
| 评阅意见： 该论文以提升水质监测效率，采用了自适应采样算法和机器学习算法，设计实现了一个智能水质监测系统。论文表述较清楚，结构较完整。不足之处在于：论文详述了数据获取与采用的模型方法，但对系统的分析、设计与实现部分表述略显简略；缺乏对系统的测试与分析。该论文达到本科毕业设计要求，同意该生参加论文答辩。   |  |  |  |  |  |  | | --- | --- | --- | --- | --- | --- | | 评阅人签字： | |  | | | | | 2024 | 年 | 5 | 月 | 31 | 日 | | | | | | | | | |
| 评定结论：（在相应栏划“√”） | | | 同意答辩（） | | 修改后答辩（） | | | 不同意答辩（） |

**重庆邮电大学本科毕业设计（论文）答辩记录表**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| 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 | |
| 学生陈述、答辩组提问及学生答辩情况：  学生陈述：  毕设内容主要是基于ADASYN 和机器学习模型开发了一个水质监测系统。首先数据经过ADASYN和smote算法进行过采样。然后过采样算法后的数据发送给机器学习模型进行训练机器学习模型包含KNN、SVM、GDBT及ANN，对模型的超参数进行调优，选择最优超参数，然后进行模型预测。训练好的模型，通过精确率、召回率以及F1得分指标发送给前端展示。系统前后端框架主要采用是Vue+ElementUi作为前端，后端框架是用Django实现MVC架构。  国内外研究:国外主要侧重于算法的集成，国内主要侧重于数据集的可靠性。系统的需求主要分为功能需求和非功能需求，包含文件上传模块，数据预处理模块，模型训练及预测模块，前端展示模块和后端服务模块。非功能需求包含性能方面，可用性方面，兼容性方面以及可维护性方面。数据集来源于Kaggle平台。自变量包含温度、溶氧量、pH值、电导率、生化需氧量、硝酸盐和亚硝酸盐的平均值、粪大肠杆菌，因变量是总大肠菌。技术栈包括scikit-learn算法库选择、TensorFlow深度学习框架，Django后端框架部署以及Vue+Elementui前端框架。  系统详细设计主要讲解了数据预处理模块和模型训练与预测模块。数据预处理模块包含数据清洗、特征选择和过采样处理，模块训练与预测模块对四种学习模型调参。KNN的超参数为k值；GDBT的超参数为学习率；SVM的超参数为c值；ANN的超参数为k值。计算其精确率，召回率以及F1得分，最终将结果返回前端页面展示。  答辩组提问及学生答辩情况：  问：KNN算法中k值代表什么： 答：K值代表邻居数量，表示欧几里得距离，两点之间的直线距离。 问：系统研发体现在哪里 答：本研究主要侧重于算法部分，对于系统设计方面，仅使用前后端页面进行演示，因此系统部分内容并不多，在后续会考虑增加系统设计图等内容。 问：过采样算法对后续模型训练的影响 答：过采样算法通过插值的方式生成一些数据点，这些值不会对后面模型训练产生影响，根据论文作为依据。  问：软件周期过程包含哪些  答：首先需求分析，根据需求分析生成用例图，然后进行系统的总体设计与开发 问：数据集在论文哪里 答：在论文数据获取那一栏，但是缺少对数据是否是连续变量还是离散变量以及数据量的描述，列变量表格。 问：召回率和准确率、F1指标 答：准确率是预测结果中预测正确值的比例，召回率是真实结果中预测正确的比例，F1是他们的调和平均值，其意义就是认为召回率和精准率同等重要，防止过高的准确率下，出现过低的召回率。 问：什么是过拟合  答：过拟合是指训练集结果非常好，损失函数很低，但在交叉验证集损失函数很高，没有很好的鲁棒性。 | | | | | | |
| 评分项目 | | | 满分分值 | | | 评定评分 |
| 1 | 课题工作量 | | 10 | | | 6 |
| 2 | 工作完成质量 | | 30 | | | 18 |
| 3 | 答辩准备情况 | | 10 | | | 6 |
| 4 | 答辩展示与内容陈述 | | 20 | | | 12 |
| 5 | 提问回答情况 | | 30 | | | 18 |
| 答辩小组评定成绩（百分制） | | | | | | 60 |
| 答辩小组结论：  请直接写：答辩小组同意通过/不通过毕业设计论文答辩 | | | | | | |
| |  |  |  |  |  |  | | --- | --- | --- | --- | --- | --- | | 组长签字： | |  | | | | | 2024 | 年 | 6 | 月 | 4 | 日 | | | | | | | |

**重庆邮电大学本科毕业设计（论文）答辩委员会意见及评分表**

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 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 | | |
| 指导教师评定成绩 | |  | | 评阅人评定成绩 |  | | 答辩小组评定成绩 |  | |
| 答辩委员会评定成绩（百分制） | | |  | | | | | |
| 答  辩  委  员  会  意  见 | 经过对毕业设计（论文）及其相关教学档案的审查，答辩委员会认定该生（□达到/□未达到）本科毕业设计（论文）的要求，（□同意/□不同意）通过该生论文答辩。   |  |  |  |  |  |  | | --- | --- | --- | --- | --- | --- | | 负责人签章： | |  | | | | |  | 年 |  | 月 |  | 日 | | | | | | | | |
| 最  终  等  级 | 请勾选最终成绩（单选）：  □ A □ B+ □B- □ C □ F   |  |  |  |  |  |  | | --- | --- | --- | --- | --- | --- | | 学院成绩专用章： | |  | | | | |  | 年 |  | 月 |  | 日 | | | | | | | | |