**河北工程大学**

**本科毕业设计(论文)任务书**

**题目：基于机器学习的脑卒中预测**

**学生姓名：** 袁留威

**专 业：**数据科学与大数据技术

**班级学号：** 190440227

**指导老师：** 吕艳芬

**指导教师职称：** 讲师

**2023年 1 月 10 日**

**一、题目**

基于机器学习的脑卒中预测

**二、预计成果**

使用机器学习方法来探究影响中风的主要因素并对脑卒中进行预测，选择合适的分类模型，得到影响中风的主要特征，预测一个人患中风的可能性，帮助医生决策，为大家尤其是老人群体提供预防中风的建议。

**三、目的与意义**

调查显示，城乡合计脑卒中已成为我国第一位死亡原因，也是中国成年人残疾的首要原因，脑卒中具有发病率高、死亡率高和致残率高的特点，早期发现并对其干预治疗对提升脑卒中治疗效果减少医疗花费具有重要意义。目前脑卒中病情诊断主要依赖于医生的工作经验，受到医生的主观因素影响。为此本文基于机器学习构建脑卒中预测模型，实现对患者脑卒中病情的客观筛查及预防，用以辅助医生决策。

**四、任务与要求**

任务：本次研究的主要任务是基于建立机器学习模型来探究影响中风的主要因素并对脑卒中进行预测。通过描述性统计分析得到各变量之间的相关性情况；通过分析脑卒中预测数据集，将是否患有中风作为分类目标，分析影响中风的指标；使用机器学习方法建模评估，选择合适的分类模型，得到影响中风的主要特征，用以辅助医生决策。

要求：根据研究方向，独立进行文献查找和分析文献资料；能够独立查找、翻译和分析外文资料；参考国内外研究现状和成果、独立分析、写作、完成完整的毕业设计。

1. 开题报告不少于3000字，文献综述不少于2000字；
2. 翻译文章不少于2500汉字(附原文)；
3. 正文不少于18000字，包括200字左右的中英摘要，3~5关键词。

**五、具体内容**

第一部分 使用python工具对脑卒中数据进行包括均值填充、数据变换、数据清理等的数据预处理，运用过采样方法解决数据不平衡问题。

第二部分 通过描述性统计分析得到各变量之间的相关性情况，得到影响中风的主要特征。

第三部分 阐述机器学习方法的基本理论，包括随机森林，决策树等，模型训练过程中遇到的问题以及解决方法。

第四部分 选取多个机器学习模型进行预测分析，并对模型性能评估。

第五部分 通过对各个模型指标的对比，选出最优的机器学习模型，并使用最优的模型进行预测分析，并利用预测结果辅助医生决策。

**六、研究成果**

通过描述性统计分析得到各变量之间的相关性情况，并得到影响中风的主要因素。利用机器学习对脑卒中进行预测，不断提高脑卒中预测模型的精度，实现对患者脑卒中病情的客观筛查及预防，并利用预测结果辅助医生进行决策，成果以毕业设计展示。

**七、时间安排**

|  |  |  |  |
| --- | --- | --- | --- |
| 序号 | 论文各阶段名称 | 日期（教学周） | 备注 |
| 1 | 提交毕业论文题目 | 第7学期期末（1月） | 指导老师提交选题申请表 |
| 2 | 下达毕业论文指导书 | 第7学期期末（1月） | 指导老师提交任务书 |
| 3 | 收集阅读文献，调研最新研究进展，完成开题报告和外文翻译。 | 第8学期第4周—第5周 | 学生提交开题报告、外文翻译 |
| 4 | 收集阅读文献，针对论文题目开展研究。 | 第6周—第9周 | 完成中期检查 |
| 5 | 撰写论文，完成初稿。 | 第10周—第13周 | 学生提交论文初稿 |
| 6 | 修改论文，完成第二稿。 | 第14周 | 学生提交论文第二稿 |
| 7 | 进一步修改论文，完成论文定稿。 | 第15周 | 学生提交论文定稿 |
| 8 | 完成论文装订，指导教师、评阅教师先后评阅论文。 | 第16周 | 学生装订论文，指导老师、评阅教师完成打分 |
| 9 | 毕业答辩 | 第17周 | 教师提交学生毕业论文最终成绩 |

**八、参考资料**

[1] 李洁洁,张雁儒,李昊,杨越.机器学习在脑卒中预测中的研究进展[J].河南医学研究,2022,31(20):3832-3835.

[2] 常怀文,姚音.基于机器学习构建江西地区缺血性脑卒中风险预测模型[J].西部医学,2022,34(08):1182-1186.

[3] 侯玉梅,曾慧,张晨阳,张重阳,张利军,高秋烨.基于数据挖掘的缺血性脑卒中患病风险预测[J].中国老年学杂志,2021,41(01):177-181.

[4] 武茂林. 基于机器学习预测缺血性脑卒中复发的研究[D].山西医科大学,2022.DOI:10.27288/d.cnki.gsxyu.2022.000333.

[5] 上官艺,王孟,王春娟,谷鸿秋,赵性泉,王伊龙,王拥军,李子孝.基于机器学习的缺血性卒中功能预后预测模型研究[J].中国卒中杂志,2021,16(09):895-900.

[6] 刘洋. 基于机器学习的脑卒中预测模型的研究及其应用[D].东北林业大学,2022.DOI:10.27009/d.cnki.gdblu.2022.001245.

[7] 饶夫阳,宋艳平,吕芯芮,白旭,覃伟,刘欢,曹琦,李子孝,刘宝花,姜勇.基于机器学习模型缺血性脑卒中1年死亡预测效果评价[J].中国公共卫生,2019,35(09):1187-1191.

[8] 彭晨. 基于机器学习算法的缺血性脑卒中患者预后影响因素及预测研究[D].南昌大学,2021.DOI:10.27232/d.cnki.gnchu.2021.001034.

[9] 范永东. 模型选择中的交叉验证方法综述[D].山西大学,2013.

[10] 文谦,杨家伟,武泽平,杨希祥,赵海龙,王志祥.快速交叉验证改进的运载火箭近似建模方法[J].航空学报,2022,43(09):461-474.

[11] 郭志恒,刘青萍,刘芳,王成武,阮旭凌.基于机器学习算法的脑卒中疾病早期预测模型研究[J].计算机与数字工程,2021,49(11):2180-2183+2247.

[12] Dev Soumyabrata et al. A predictive analytics approach for stroke prediction using machine learning and neural networks[J]. Healthcare Analytics, 2022, 2

[13] Machine Learning; Researchers from Xiamen University Describe Findings in Machine Learning (Stroke Prediction with Machine Learning Methods among Older Chinese)[J]. Robotics & Machine Learning, 2020,

[14] Chen Yu Chen et al. Building a Classifier of Onset Stroke Prediction Using Random Tree Algorithm[J]. International Journal of Machine Learning and Computing, 2017, 7(4) : 61-66.