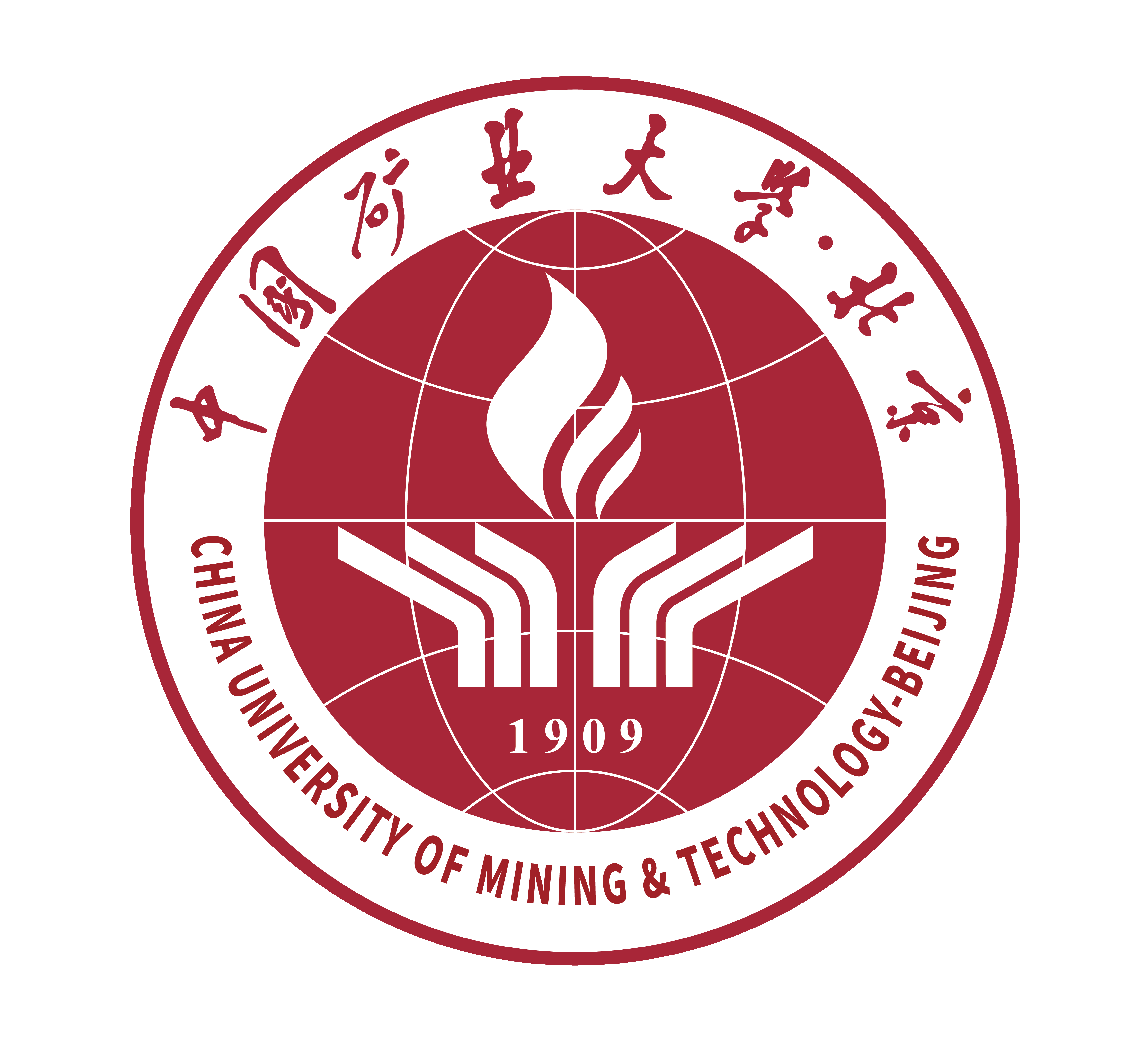
China University of Mining & Technology-Beijing



**本科生毕业设计(论文)**

基于模型融合的在线招聘欺诈检测

|  |  |  |
| --- | --- | --- |
| 学 院 |  | 管理学院 |
| 专业班级 |  | 大数据2021-2班 |
| 学号 |  | 2110570430 |
| 姓 名 |  | 周伟诺 |
| 指导教师 |  | 李岩 |

2025年6月

**中国矿业大学（北京）**

**本科生毕业设计（论文）**

|  |  |  |  |
| --- | --- | --- | --- |
| 中文题目： | 基于模型融合的在线招聘欺诈检测 | | |
| 英文题目： | Detecting Online Recruitment Fraud Using Mode Fusion Techniques | | |
| 姓 名： | 周伟诺 | 学 号： | 2110570430 |
| 学 院： | 管理学院 | | |
| 专 业： | 大数据管理与应用 | 班 级： | 大数据管理与应用2021-2班 |
| 指导教师： | 李岩 | 职 称： | 副教授 |
| 完成日期： | 2025 年 月 日 | | |

诚信声明

本人郑重声明：所呈交的毕业设计（论文）是本人在指导教师的指导下独立完成的。除文中已经注明引用的内容外，毕业设计（论文）中不包含其他人已经发表或撰写过的研究成果。对本文做出重要贡献的个人和集体，均已在文中以明确方式标明。

作者签名： 日期：

关于使用授权的说明

本人完全了解中国矿业大学（北京）有关保留、使用毕业设计（论文）的规定，即：学校有权保留送交论文的复印件，允许论文被查阅或借阅；学校可以公布论文的全部或部分内容，可以采用影印、缩印或其他复制手段保存论文。

作者签名： 导师签名： 日期：

**中国矿业大学（北京）本科生毕业设计（论文）任务书**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| 学院： |  | | 专业： |  | 班级： |  |
| 姓名： |  | | 学号： |  | | |
| 任务下达日期： | | 20 年 月 日 | | | | |
| 任务完成日期： | | 20 年 月 日 | | | | |
| 论文题目： | |  | | | | |
| 专题题目： | |  | | | | |
| **任务主要内容：** | | | | | | |
|  | | | | | | |
| **任务目标要求**（文献阅读、外文资料翻译、设计或实验工作量，图纸、软硬件数量及技术指标等）： | | | | | | |
|  | | | | | | |
| **时间进度安排：** | | | | | | |
|  | | | | | | |
| **推荐阅读的文献资料：** | | | | | | |
|  | | | | | | |
| **教学院长签字：** | |  | | **指导教师签字：** |  | |

**中国矿业大学（北京）**

**本科生毕业设计（论文）结合科研说明书**

学院：

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 学生姓名 |  | 专业 |  | | | 班级 | | |  |
| 题目名称 |  | | | | | | | | |
| 题目种类 |  | 题目类型 |  | | | | | | |
| 指导教师 |  | 专业 |  | | | 职称 | |  | |
| 科研课题  基本情况 | 科研课题名称 |  | | | | | | | |
| 科研课题来源 |  | | | 科研立项  起止时间 | | 20 年至20 年 | | |
| 主要研究内容: | | | | | | | | |
| 学生参与科研课题研究情况 | 参与课题研究的内容： | | | | | | | | |
| 参与研究的工作量: | | | | | | | | |
| 系（教研室）意见 | | | | 学院意见 | | | | | |
| 毕业设计（论文）内容与科研课题相关，并且学生参与科研课题研究，同意认定毕业设计（论文）结合科研课题。  系（教研室）主任签字：  20 年 月 日 | | | | 毕业设计（论文）结合科研课题，并且符合学生创新学分认定要求，同意给予学生毕业设计（论文）结合科研创新学分。  主管院长签字：  20 年 月 日 | | | | | |

**摘 要**

运用自制的可视化实验设备，进行了表面活性剂吐温（T40，1×10-3 mol/L、T40，2×10-3 mol/L、T40/T80（1:1），1×10-3 mol/L）在瓦斯水合过程中影响研究，结合煤层高压注水中添加表面活性剂R1-89的现场研究，分析了表面活性剂在瓦斯水合三阶段的主要作用。认为在煤层注水阶段表面活性剂的加入降低了液体表面张力和注水压力，使注水速度加快，煤体得到均匀润湿；在水合物诱导阶段表面活性剂的胶束化增加了瓦斯气体的溶解度，促进了气体在溶液中过饱和，推动了水合物晶核生长，缩短了诱导时间；在水合物生长阶段表面活性剂胶束对溶于其中的气体分子和吸附于其周围的水分子的束缚作用，相当于降低了体系的温度，改变了水合物生成热力学条件。

（摘要要求用精炼的文字对论文的目的、内容、方法、成果、结论等进行高度概括，不含公式、图表，不加注释，去掉“本文”等第1人称字样，不出现参考文献序号，具有独立性和完整性。中文摘要一般为400字左右。）

**关键词：**瓦斯含量；突出预测；敏感性；预测深度

**ABSTRACT**

Using self-developed visible experimental installation, the influences of surfactant Tween (T40, 1×10-3mol/L; T40, 2×10-3mol/L; T40/T80(1:1), 1×10-3mol/L) on the process of gas hydration were studied. Combined with the field study of adding surfactant R1-89 for coal seam water infusion with high pressure, the essential roles of surfactant in three stages of gas hydration were analyzed. The results indicate that in the stage of coal seam water infusion, the surfactant decreases the surface tension of liquid and the pressure of water infusion, quickens the rate of water infusion, and coal is dabbled uniformly; in the stage of gas hydrate induction, the micelle agglomeration of surfactant can increase the solubility of gas, accelerate the supersaturation of gas in solution, prompt the formation of crystal nucleus of gas hydrate, and the induction time is shortened; in the stage of gas hydrate growth, the binding effect of surfactant micelle on gas molecule dissolved in it and water molecule absorbed around it is equivalent to the reduction of systems’ temperature, so the thermodynamic conditions of gas hydrate formation are changed.

（英文摘要应尽量与中文相对应。）

**Keywords：**gas content; outburst danger prediction; sensitivity; prediction

目 录

[1 绪论 1](#_Toc191109207)

[1.1. 研究背景与意义 1](#_Toc191109208)

[1.1.1. 在线招聘行业现状 1](#_Toc191109209)

[1.1.2. 招聘欺诈问题的严重性 1](#_Toc191109210)

[1.1.3. 研究的目的与意义 1](#_Toc191109211)

[1.2. 国内外研究现状 1](#_Toc191109212)

[1.2.1. 招聘欺诈检测研究现状 1](#_Toc191109213)

[1.2.2. 多模型融合方法研究现状 1](#_Toc191109214)

[1.2.3. 研究现状总结与分析 1](#_Toc191109215)

[1.3. 研究内容与方法 1](#_Toc191109216)

[1.3.1. 研究内容概述 1](#_Toc191109217)

[1.3.2. 研究方法与技术路线图 1](#_Toc191109218)

[1.4. 论文组织结构安排 1](#_Toc191109219)

[2 相关理论与技术基础 1](#_Toc191109220)

[2.1. 在线招聘欺诈相关概述 1](#_Toc191109221)

[2.1.1. 在线招聘欺诈的定义与类型 1](#_Toc191109222)

[2.1.2. 在线招聘欺诈的特征与识别难点 1](#_Toc191109223)

[2.2. 自然语言处理 2](#_Toc191109224)

[2.2.1. 自然语言处理在在线招聘欺诈检测中的应用 2](#_Toc191109225)

[2.2.2. 常用的自然语言处理工具 2](#_Toc191109226)

[2.3. 特征工程 2](#_Toc191109227)

[2.3.1. 特征工程的概念与重要性 2](#_Toc191109228)

[2.3.2. 特征转换 2](#_Toc191109229)

[2.4. 机器学习基础模型 2](#_Toc191109230)

[2.4.1. 逻辑回归模型 2](#_Toc191109231)

[2.4.2. 决策树模型 2](#_Toc191109232)

[2.4.3. 支持向量机模型 2](#_Toc191109233)

[2.4.4. 随机森林模型 2](#_Toc191109234)

[2.4.5. 神经网络模型 2](#_Toc191109235)

[2.5. 多模型融合方法 2](#_Toc191109236)

[2.5.1. 投票法 2](#_Toc191109237)

[2.5.2. 均值法 2](#_Toc191109238)

[2.5.3. Stacking法 2](#_Toc191109239)

[2.5.4. Blending法 2](#_Toc191109240)

[2.6. 本章小结 2](#_Toc191109241)

[3 在线招聘数据收集与预处理 3](#_Toc191109242)

[3.1. 数据收集 3](#_Toc191109243)

[3.1.1. 数据来源 3](#_Toc191109244)

[3.1.2. 公开数据集介绍 3](#_Toc191109245)

[3.2. 数据集描述 3](#_Toc191109246)

[3.2.1. 数据集的规模与特征 3](#_Toc191109247)

[3.2.2. 数据的标签信息 3](#_Toc191109248)

[3.3. 数据预处理 3](#_Toc191109249)

[3.3.1. 数据合并 3](#_Toc191109250)

[3.3.2. 重复值处理 3](#_Toc191109251)

[3.3.3. 缺失值处理 3](#_Toc191109252)

[3.3.4. 异常值处理 3](#_Toc191109253)

[3.4. 本章小结 3](#_Toc191109254)

[4 特征工程 3](#_Toc191109255)

[4.1. 特征构造 3](#_Toc191109256)

[4.1.1. 语义特征构造 3](#_Toc191109257)

[4.1.2. 元特征构造 3](#_Toc191109258)

[4.2. 特征转换 3](#_Toc191109259)

[4.2.1. 有序分类特征转换 4](#_Toc191109260)

[4.2.2. 基于频率的特征转换 4](#_Toc191109261)

[4.2.3. 留一编码 4](#_Toc191109262)

[4.2.4. 数据标准化 4](#_Toc191109263)

[4.2.5. TF-IDF向量化 4](#_Toc191109264)

[4.3. 本章小结 4](#_Toc191109265)

[5 基础分类模型构建与训练 4](#_Toc191109266)

[5.1. 模型构建 4](#_Toc191109267)

[5.1.1. 逻辑回归模型 4](#_Toc191109268)

[5.1.2. 决策树模型 4](#_Toc191109269)

[5.1.3. 随机森林模型 4](#_Toc191109270)

[5.1.4. 支持向量机模型 4](#_Toc191109271)

[5.1.5. 神经网络模型 4](#_Toc191109272)

[5.2. 模型参数设置及训练 4](#_Toc191109273)

[5.2.1. 训练集及测试集划分 4](#_Toc191109274)

[5.2.2. 模型训练 4](#_Toc191109275)

[5.3. 模型评估 4](#_Toc191109276)

[5.3.1. 评估指标选择 4](#_Toc191109277)

[5.3.2. 各个基础分类模型评估结果分析 4](#_Toc191109278)

[5.4. 本章小结 5](#_Toc191109279)

[6 多模型融合策略与实现 5](#_Toc191109280)

[6.1. 多模型融合框架设计 5](#_Toc191109281)

[6.1.1. 融合的目标与原则 5](#_Toc191109282)

[6.1.2. 融合框架总体架构 5](#_Toc191109283)

[6.2. 不同融合方法的实现 5](#_Toc191109284)

[6.2.1. 投票法 5](#_Toc191109285)

[6.2.2. 均值法 5](#_Toc191109286)

[6.2.3. Stacking法 5](#_Toc191109287)

[6.2.4. Blending法 5](#_Toc191109288)

[6.3. 融合模型的评估与比较 5](#_Toc191109289)

[6.3.1. 融合模型评估指标 5](#_Toc191109290)

[6.3.2. 不同融合模型的性能对比 5](#_Toc191109291)

[6.4. 本章小结 5](#_Toc191109292)

[7 模型部署与应用 5](#_Toc191109293)

[7.1. 模型部署方案 5](#_Toc191109294)

[7.1.1. 部署环境选择 5](#_Toc191109295)

[7.1.2. 部署流程与步骤 5](#_Toc191109296)

[7.2. 系统实现与测试 5](#_Toc191109297)

[7.2.1. 在线招聘欺诈检测系统设计 6](#_Toc191109298)

[7.2.2. 系统功能测试与验证 6](#_Toc191109299)

[7.3. 实际应用效果分析 6](#_Toc191109300)

[7.3.1. 应用场景与数据 6](#_Toc191109301)

[7.3.2. 实际效果评估 6](#_Toc191109302)

[7.4. 本章小结 6](#_Toc191109303)

[8 总结与展望 6](#_Toc191109304)

[8.1. 研究成果总结 6](#_Toc191109305)

[8.1.1. 主要研究内容回顾 6](#_Toc191109306)

[8.1.2. 研究成果与创新点 6](#_Toc191109307)

[8.2. 研究不足与展望 6](#_Toc191109308)

[8.2.1. 研究过程中的不足 6](#_Toc191109309)

[8.2.2. 未来研究方向与建议 6](#_Toc191109310)

[参考文献 7](#_Toc191109311)

[附录1 代码清单 12](#_Toc191109312)

[附录2 数据集详细说明及分析 13](#_Toc191109313)

[附录3 实验相关补充材料 14](#_Toc191109314)

插图索引

[图 1‑1：2022年中国网络招聘行业市场规模 1](#_Toc191202718)

[图 1‑2：招聘平台及形式多样化 2](#_Toc191202719)

[图 1‑3：研究技术路线图 6](#_Toc191202720)

表格索引

[表格 1-1：测试表格 4](#_Toc191202500)

[表格 2-1：测试表格2 4](#_Toc191202501)

# 绪论

* 1. 研究背景与意义

在数字化时代的浪潮下，互联网技术的飞速发展深刻地改变了人们的生活和工作方式，在线招聘行业应运而生并取得了长足的进步。随着智能手机的普及和网络覆盖范围的扩大，越来越多的求职者和招聘者倾向于通过在线招聘平台来满足各自的需求。

* + 1. 在线招聘行业现状

近年来，互联网的快速发展深刻改变了人力资源市场的招聘模式。传统的线下招聘形式逐渐被新型的招聘方式所取代 。尤其是在互联网技术高度普及的背景下，网络招聘行业如雨后春笋般涌现，成为企业招聘和求职者求职的重要桥梁。



图 ‑1：2022年中国网络招聘行业市场规模

这种变革不仅体现在招聘平台的多样化上，还体现在技术手段的创新和应用上。随着云计算、大数据、人工智能（Artificial Intelligence，AI）等前沿技术的不断成熟，招聘形式也逐渐朝着智能化和数字化的方向发展。例如，云招聘的兴起，让企业能够通过线上平台发布招聘信息、筛选简历，并组织视频面试，从而降低了传统招聘过程中时间和地域的限制。此外，社交媒体招聘也逐渐成为一种新的趋势，通过社交平台如小红书、LinkedIn和脉脉，以及短视频平台如抖音、快手和微信视频号等等，企业不仅能广泛宣传招聘信息，还能直接与潜在候选人互动，从而提升招聘效果。同时，人工智能技术的应用进一步提升了招聘的智能化水平，基于AI的招聘系统可以实现简历的自动筛选、候选人匹配以及智能推荐，有效减少了人力资源部门的重复性工作，提高了整体招聘效率。



图 ‑2：招聘平台及形式多样化

然而，网络招聘行业在迅速发展的同时，虚假招聘信息也逐渐泛滥，严重影响了在线招聘市场的健康发展。

* + 1. 招聘欺诈问题的严重性

根据Greenhouse的研究，约18%到22%的在线招聘广告实际上并没有招聘意图，这些被称为“幽灵职位”的广告不仅对求职者造成困扰，还导致了对潜在雇主的不信任感。虚假招聘信息通常通过虚构职位、夸大薪资待遇、伪装成知名企业等手段诱导求职者上当，最终通过获取个人信息、收取押金或培训费用等方式牟取非法利益。这些虚假招聘信息形式多样，不仅让求职者的时间和精力付之东流，还可能导致其个人隐私泄露、经济损失，甚至身陷诈骗陷阱。虚假招聘信息的出现，给社会带来了诸多负面影响。一方面，求职者在面对大量虚假招聘广告时，常常无法有效辨别真假，导致个人信息泄露、金钱被骗，甚至一些人因高薪诱惑而误入非法传销组织。另一方面，虚假招聘信息的泛滥严重影响了求职者对网络招聘平台的信任，阻碍了招聘生态的健康发展。企业则面临虚假招聘广告冒用其名义发布信息的风险，这不仅损害了企业形象，还可能导致求职者对该企业产生误解，进而影响其品牌声誉。

* + 1. 研究目的与意义
       1. 研究目的

目前，尽管部分网络招聘平台已建立了举报和人工审核机制，但由于虚假招聘信息的数量庞大、发布渠道复杂，传统的人工审核方法不仅效率低下，而且难以覆盖所有的虚假信息。尤其随着招聘广告形式日益多样化，人工审核的局限性愈发显现。例如，仅用基于规则的模型，难以应对表述越来越灵活的虚假信息，无法在实际应用中很好地满足快速、精准识别虚假招聘信息的需求。因此，如何通过技术手段实现对虚假招聘信息的快速识别，已成为招聘行业亟待解决的难题。

* + - 1. 研究意义

1. 保护求职者的利益：求职者是虚假招聘信息的主要受害者。虚假招聘信息通常通过夸大的薪资待遇、虚构的职位要求等手段诱骗求职者，造成时间、金钱和个人隐私的损失。本论文旨在通过基于多模型融合的虚假招聘信息识别，可以有效避免求职者受到欺诈，保护他们的经济利益和个人安全。
2. 推动招聘行业的健康发展：虚假招聘信息不仅影响求职者和平台，也可能扰乱招聘市场的正常秩序。本论文通过提高虚假信息识别的效率和准确性，有助于净化招聘市场，促进行业的健康可持续发展，降低招聘平台的运营风险。
3. 助力监管与平台管理：监管部门在打击虚假招聘乱象时，需要有效的技术手段来辅助判断信息的真实性。多模型融合的识别方法可为监管部门提供可靠的数据支撑和判断依据，使其能更精准地发现并查处发布虚假招聘信息的主体，增强监管的有效性，维护良好的市场监管秩序。
   1. 国内外研究现状
      1. 招聘欺诈检测研究现状
      2. 多模型融合方法研究现状
      3. 研究现状总结与分析
   2. 研究内容与方法
      1. 研究内容概述
      2. 研究方法与技术路线图

本研究的研究方法主要有四种：

1. **实证分析法**：实证分析法是深入探究在线招聘信息检测问题的重要手段。首先，明确定义一系列与研究紧密相关的核心变量。这些变量涵盖了招聘信息的多个维度，例如招聘职位的描述详细程度、薪资待遇的合理性范围、公司信息的完整性等。通过对这些变量的精确界定，为后续的研究提供了清晰且可量化的研究对象。基于这些核心变量，精心构建相应的模型。此模型旨在揭示各变量之间的内在关系以及它们对招聘信息真实性判断的影响机制。在构建过程中，充分考虑到招聘领域的特点和实际情况，确保模型具有较高的针对性和实用性。运用多种统计分析方法开展实证研究。描述性统计分析用于对收集到的数据进行初步整理和概括，获取各个变量的基本特征，如均值、中位数、标准差等。这些统计量能够帮助我们快速了解数据的集中趋势、离散程度以及分布形态，为进一步的分析提供基础。相关性统计分析则聚焦于探究变量之间的关联程度，通过计算相关系数，明确不同变量之间是正相关、负相关还是无明显关联。这种分析有助于我们找出对虚假招聘信息识别具有关键作用的变量，为后续的模型优化和解释提供有力支持。通过实证分析法，我们能够以数据为依据，深入剖析招聘信息中的内在规律，为虚假招聘信息的检测提供科学、可靠的理论依据。
2. **文本挖掘法**：鉴于招聘信息主要以文本形式呈现，文本挖掘法在本研究中发挥着关键作用。针对研究内容的特点和需求，首先对收集到的数据进行全面细致的清理，去除其中的噪声信息。这些噪声可能包括格式错误、乱码、无关的特殊字符以及重复冗余的内容等。通过数据清理，能够提高数据质量，为后续的特征提取提供纯净、有效的数据基础。在清理后的数据基础上，运用先进的自然语言处理技术，从招聘信息中深度提取与虚假信息识别紧密相关的语义特征。这些语义特征涵盖多个层面，例如词汇层面，通过分析招聘信息中特定词汇的出现频率、词向量等，挖掘出与虚假信息相关的高频词汇或具有特殊语义指向的词汇；句法层面，研究句子结构、语法规则以及词性搭配等方面的特征，识别出不符合正常招聘信息表达习惯的句式；语义层面，借助语义角色标注、主题模型等技术，理解招聘信息的深层语义，发现潜在的语义矛盾或不合理之处。通过文本挖掘法提取的语义特征，能够从文本的本质层面为虚假招聘信息的识别提供丰富、准确的线索，有助于提高识别模型的准确性和有效性。
3. **实验研究法**：构建多个不同类型的基础分类模型用于招聘信息检测，如逻辑回归模型、决策树模型、随机森林模型、支持向量机模型、神经网络模型等。针对不同模型设置不同的参数组合，开展多组实验，对比各模型在相同数据集上的性能表现，包括准确率、召回率、F1 值等指标。
4. **模型融合法**：单一模型在面对复杂多变的在线招聘信息时，往往存在一定的局限性，难以全面、准确地识别虚假招聘信息。为了克服这一问题，本研究采用多模型融合法，旨在通过整合多个模型的优势，提高虚假招聘信息识别的准确性和鲁棒性。多模型融合法的核心思想是将不同类型基础模型在处理招聘信息时所捕捉到的信息和特征进行有机结合。这些基础模型由于各自的算法原理和学习方式不同，对招聘信息的理解和表征也有所差异。例如，逻辑回归模型擅长捕捉线性关系，决策树模型能够直观地展示数据的分类边界，随机森林模型通过集成多个决策树提高了稳定性，支持向量机模型在处理非线性问题时表现优异，神经网络模型则能够自动学习深层次的语义特征。在多模型融合过程中，采用多种融合策略。常见的融合策略包括投票法、加权平均法、Stacking法和Blending法等。投票法是最简单直接的融合方式，各个基础模型根据自己的预测结果进行投票，最终选择得票最多的类别作为融合模型的预测结果。加权平均法在平均法的基础上，为不同的基础模型分配不同的权重，权重的确定通常基于模型在训练集上的性能表现，性能较好的模型赋予较高的权重，从而更充分地发挥优势模型的作用。Stacking 法是一种更为复杂的融合策略，它使用一个新的模型（元模型）来学习如何组合各个基础模型的预测结果，通过将基础模型的预测结果作为元模型的输入，训练元模型以获得更优的融合效果。通过多模型融合法，能够充分整合各个基础模型的优势信息，弥补单一模型的不足，有效提高虚假招聘信息识别的准确性和鲁棒性。同时，面对招聘信息数据的动态变化和噪声干扰，融合模型能够展现出更强的适应性和稳定性，为实际应用提供更可靠的虚假招聘信息识别解决方案。

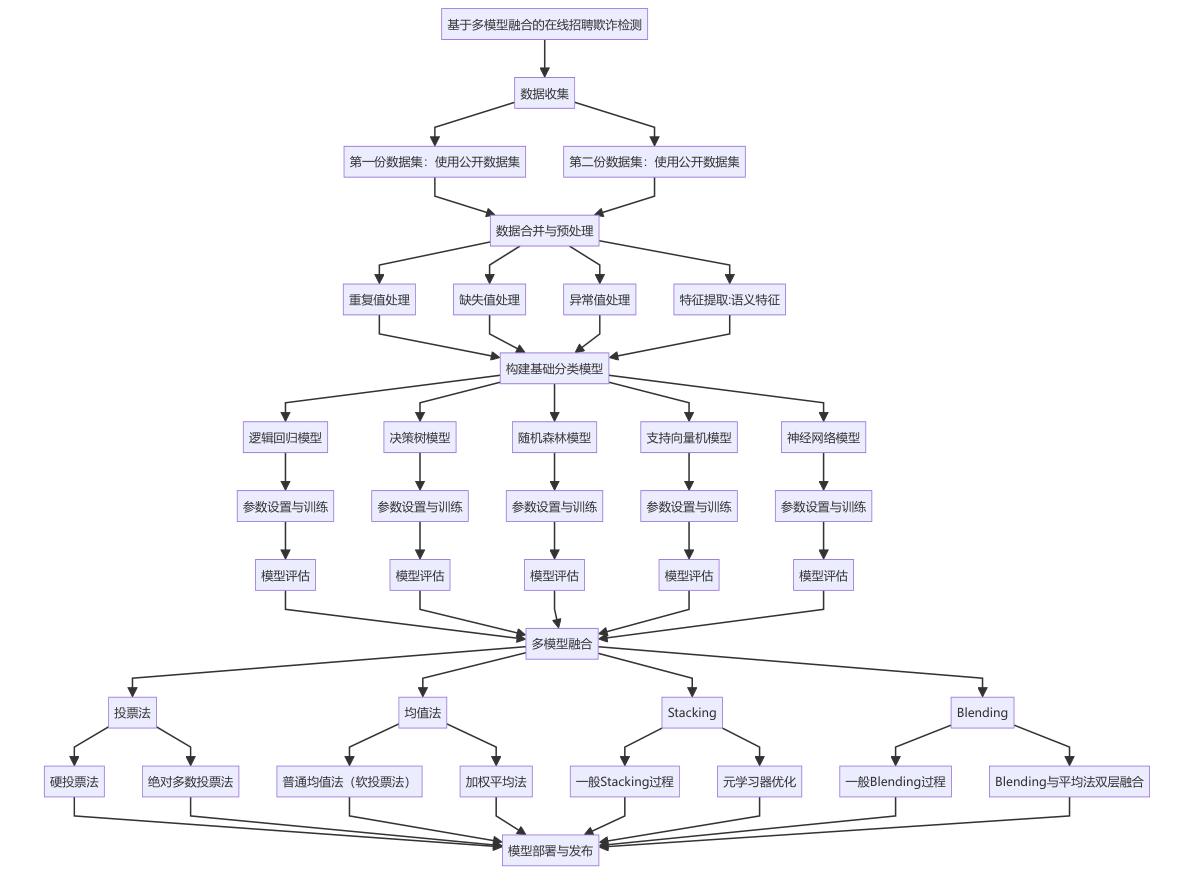


图 ‑3：研究技术路线图

* 1. 论文组织结构安排

# 相关理论与技术基础

* 1. 在线招聘欺诈相关概述
     1. 在线招聘欺诈的定义与类型
     2. 在线招聘欺诈的特征与识别难点

表格 2-1：测试表格2

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |

* 1. 自然语言处理
     1. 自然语言处理在在线招聘欺诈检测中的应用
     2. 常用的自然语言处理工具
  2. 特征工程
     1. 特征工程的概念与重要性
     2. 特征转换
  3. 机器学习基础模型
     1. 逻辑回归模型
     2. 决策树模型
     3. 支持向量机模型
     4. 随机森林模型
     5. 神经网络模型
  4. 多模型融合方法
     1. 投票法
     2. 均值法
     3. Stacking法
     4. Blending法
  5. 本章小结

# 在线招聘数据收集与预处理

* 1. 数据收集
     1. 数据来源
     2. 公开数据集介绍
  2. 数据集描述
     1. 数据集的规模与特征
     2. 数据的标签信息
  3. 数据预处理
     1. 数据合并
     2. 重复值处理
     3. 缺失值处理
     4. 异常值处理
  4. 本章小结

# 特征工程

* 1. 特征构造
     1. 语义特征构造
     2. 元特征构造
  2. 特征转换
     1. 有序分类特征转换
     2. 基于频率的特征转换
     3. CatBoost编码
     4. 数据标准化
     5. TF-IDF向量化
  3. 本章小结

# 基础分类模型构建与训练

* 1. 模型构建
     1. 逻辑回归模型
     2. 决策树模型
     3. 随机森林模型
     4. 支持向量机模型
     5. 神经网络模型
  2. 模型参数设置及训练
     1. 训练集及测试集划分
     2. 模型训练
  3. 模型评估
     1. 评估指标选择
     2. 各个基础分类模型评估结果分析
  4. 本章小结

# 多模型融合策略与实现

* 1. 多模型融合框架设计
     1. 融合的目标与原则
     2. 融合框架总体架构
  2. 不同融合方法的实现
     1. 投票法
     2. 均值法
     3. Stacking法
     4. Blending法
  3. 融合模型的评估与比较
     1. 融合模型评估指标
     2. 不同融合模型的性能对比
  4. 本章小结

# 模型部署与应用

* 1. 模型部署方案
     1. 部署环境选择
     2. 部署流程与步骤
  2. 系统实现与测试
     1. 在线招聘欺诈检测系统设计
     2. 系统功能测试与验证
  3. 实际应用效果分析
     1. 应用场景与数据
     2. 实际效果评估
  4. 本章小结

# 总结与展望

* 1. 研究成果总结
     1. 主要研究内容回顾
     2. 研究成果与创新点
  2. 研究不足与展望
     1. 研究过程中的不足
     2. 未来研究方向与建议

参考文献

1. 刘谢彪. (2020). 互联网虚假招聘的识别与防范. 投资与合作.
2. 王瑞琪, 纪淑娟, 曹宁, & 郭亚杰. (2023). 基于一致性训练的半监督虚假招聘广告检测模型. 计算机应用, *43*(9), 2932.
3. 谢宁宁 & 杨新凯. (不详). 融合语义特征和统计特征的虚假招聘检测模型. 计算机与数字工程, *51*(10), 2379–2383.
4. .A P, D., .S, S., & .R, G. (2021). Identifying Real and Fake Job Posting-Machine Learning Approach. *IARJSET*, *8*(8). <https://doi.org/10.17148/IARJSET.2021.8857>
5. Afzal, H., Rustam, F., Aljedaani, W., Siddique, M. A., Ullah, S., & Ashraf, I. (2023). Identifying fake job posting using selective features and resampling techniques. *Multimedia Tools and Applications*, *83*(6), 15591–15615. <https://doi.org/10.1007/s11042-023-15173-8>
6. Akram, N., Irfan, R., Al-Shamayleh, A. S., Kousar, A., Qaddos, A., Imran, M., & Akhunzada, A. (2024). Online Recruitment Fraud (ORF) Detection Using Deep Learning Approaches. *IEEE Access*. <https://ieeexplore.ieee.org/abstract/document/10614582/>
7. Alandjani, G. O. (2022). Online fake job advertisement recognition and classification using machine learning. *3 c TIC: cuadernos de desarrollo aplicados a las TIC*, *11*(1), 251–267.
8. Amaar, A., Aljedaani, W., Rustam, F., Ullah, S., Rupapara, V., & Ludi, S. (2022). Detection of Fake Job Postings by Utilizing Machine Learning and Natural Language Processing Approaches. *Neural Processing Letters*, *54*(3), 2219–2247. <https://doi.org/10.1007/s11063-021-10727-z>
9. Anita, C. S., Nagarajan, P., Sairam, G. A., Ganesh, P., & Deepakkumar, G. (2021). Fake job detection and analysis using machine learning and deep learning algorithms. *Revista Geintec-Gestao Inovacao e Tecnologias*, *11*(2), 642–650.
10. Blanzieri, E., & Bryl, A. (2008). A survey of learning-based techniques of email spam filtering. *Artificial Intelligence Review*, *29*(1), 63–92. <https://doi.org/10.1007/s10462-009-9109-6>
11. Cheekati Srikanth, M. R., Ramu, S., & Guddeti, R. M. R. (2023). A Novel Fake Job Posting Detection: An Empirical Study and Performance Evaluation Using ML and Ensemble Techniques. *Security, Privacy and Data Analytics: Select Proceedings of the 2nd International Conference, ISPDA 2022*, *1049*, 219. <https://books.google.com/books?hl=zh-CN&lr=&id=V7fREAAAQBAJ&oi=fnd&pg=PA219&dq=fake+job+posting+detection&ots=81vt9-KizO&sig=TRnzhP5YkVI4Xg1trGLVfc9uu_E>
12. Chiraratanasopha, B., & Chay-intr, T. (2022). *Detecting fraud job recruitment using features reflecting from real-world knowledge of fraud*. <https://wb.yru.ac.th/bitstream/yru/6828/1/Detecting%20Fraud%20Job%20Recruitment%20Using%20Features%20Reflecting%20from%20Real-world%20Knowledge%20of%20Fraud.pdf>
13. Choudhury, D., & Acharjee, T. (2023). A novel approach to fake news detection in social networks using genetic algorithm applying machine learning classifiers. *Multimedia Tools and Applications*, *82*(6), 9029–9045. <https://doi.org/10.1007/s11042-022-12788-1>
14. Dutta, S., & Bandyopadhyay, S. K. (2020). Fake job recruitment detection using machine learning approach. *International Journal of Engineering Trends and Technology*, *68*(4), 48–53.
15. Guzella, T. S., & Caminhas, W. M. (2009). A review of machine learning approaches to spam filtering. *Expert Systems with Applications*, *36*(7), 10206–10222.
16. Habiba, S. U., Islam, M. K., & Tasnim, F. (2021). A comparative study on fake job post prediction using different data mining techniques. *2021 2nd international conference on robotics, electrical and signal processing techniques (ICREST)*, 543–546. <https://ieeexplore.ieee.org/abstract/document/9331230/>
17. Heydari, A., ali Tavakoli, M., Salim, N., & Heydari, Z. (2015). Detection of review spam: A survey. *Expert Systems with Applications*, *42*(7), 3634–3642.
18. Jagadeesh, C., Kshirsagar, P. R., Sarayu, G., Gouthami, G., & Manasa, B. (2021). Artificial intelligence based Fake Job Recruitment Detection Using Machine Learning Approach. *Journal of Engineering Sciences*, *12*, 0377–9254.
19. Jaison, S. S., & Kodabagi, M. (2023). Identifying Real and Fake Job Posting using Machine Learning. *Journal of Advanced Zoology*, *44*. <https://search.ebscohost.com/login.aspx?direct=true&profile=ehost&scope=site&authtype=crawler&jrnl=02537214&AN=174368130&h=OY%2Fayp81qvYy1iRD76KrxcyJE25fLvtUcFbaIRAEO%2Fm4X6fgg2y0%2BCGNiRZUXAMfqGO7WLL8VmpcOuDvloaL4g%3D%3D&crl=c>
20. Keerthana, B., Reddy, A. R., & Tiwari, A. (2021). Accurate Prediction of Fake Job Offers Using Machine Learning. 收入 D. Bhattacharyya & N. Thirupathi Rao (编), *Machine Intelligence and Soft Computing* (卷 1280, 页 101–112). Springer Singapore. <https://doi.org/10.1007/978-981-15-9516-5_9>
21. Khandagale, P., Utekar, A., Dhonde, A., & Karve, S. S. (2022). Fake Job Detection Using Machine Learning. *International Journal for Research in Applied Science and Engineering Technology*, *10*(4), 1822–1827.
22. Kodandaramu, C., Pillai, K. K. P., Lokesh, G., Kishore, K. M., & Kumar, A. S. (2023). *FAKE JOB POST PREDICTION USING MACHINE LEARNING*. *14*(04).
23. Mahbub, S., Pardede, E., & Kayes, A. S. M. (2022). Online recruitment fraud detection: A study on contextual features in australian job industries. *IEEE Access*, *10*, 82776–82787.
24. Mehboob, A., & Malik, M. S. I. (2021). Smart Fraud Detection Framework for Job Recruitments. *Arabian Journal for Science and Engineering*, *46*(4), 3067–3078. <https://doi.org/10.1007/s13369-020-04998-2>
25. Nandini, T., Chandrika, S. G., Mounika, P., & Kumar, V. S. (2023). Developing a model to detect fraudulent job postings: Fake vs. real. *Int J Recent Develop Sci Technol*, *7*, 52–59.
26. Naudé, M., Adebayo, K. J., & Nanda, R. (2023). A machine learning approach to detecting fraudulent job types. *AI & SOCIETY*, *38*(2), 1013–1024. <https://doi.org/10.1007/s00146-022-01469-0>
27. Nazirova, S. (2011). Survey on spam filtering techniques. *Communications and Network*, *3*(3), 153–160.
28. Palash, M. I. A., Diganto, A., Fatan, O. N., Taher, K. A., & Al Nahian, M. J. (2022). Detection of fake job postings on online using convolutional neural network. *Journal of FST*, *1*(01), 207.
29. Pillai, A. S. (2023). Detecting Fake Job Postings Using Bidirectional LSTM. *International Research Journal of Modernization in Engineering Technology and Science*. <https://doi.org/10.56726/IRJMETS35202>
30. Prashanth, C., Chandrasekaran, D., Pandian, B., Duraipandian, K., Chen, T., & Sathiyanarayanan, M. (2022). Reveal: Online fake job advert detection application using machine learning. *2022 IEEE Delhi Section Conference (DELCON)*, 1–6. <https://ieeexplore.ieee.org/abstract/document/9752784/>
31. Rajani, D., Praveena, K., Rachana, M. P., & Supriya, A. S. (2024). FAKE JOB DETECTION USING MACHINE LEARNING. *MATERIAL SCIENCE*, *23*(04). <https://materialsciencetech.com/mst/uploads/2024-42440.pdf>
32. Reddy, D. Y. V., Neeraj, B. S., Reddy, K. P., & Reddy, P. B. (2023). *ONLINE FAKE JOB ADVERT DETECTION APPLICATION USING MACHINE LEARNING*. *14*(03).
33. Rofik, R., Hakim, R. A., Unjung, J., Prasetiyo, B., & Muslim, M. A. (2024). Optimization of SVM and Gradient Boosting Models Using GridSearchCV in Detecting Fake Job Postings. *MATRIK: Jurnal Manajemen, Teknik Informatika dan Rekayasa Komputer*, *23*(2), 419–430.
34. Safeguarding Job Seekers: Research Insights into Fake Job Detection with SGD Classifier and Naive Bayes. (2024). *ResearchGate*. <https://doi.org/10.55248/gengpi.5.0324.0761>
35. Salloum, S., Tahat, K., Mansoori, A., Alfaisal, R., & Tahat, D. (2024). Analysis of Fraudulent Job Postings Using Machine Learning. *2024 International Conference on Intelligent Computing, Communication, Networking and Services (ICCNS)*, 268–270. <https://ieeexplore.ieee.org/abstract/document/10776527/>
36. Shibly, F. H. A., Uzzal, S., & Naleer, H. M. M. (2021). *Performance comparison of two class boosted decision tree snd two class decision forest algorithms in predicting fake job postings*. <http://ir.lib.seu.ac.lk/handle/123456789/5611>
37. Singh, V. R., Sampras, P., & Dhage, A. (2023). Fake Job Post Prediction Using Data Mining. *Journal of Scientific Research and Technology*, 39–47.
38. Swetha, K., Reddy, M. T., Sravani, K., Subramanyam, B., & Kumar, S. A. (2023). Fake job detection using machine learning approach. *Journal of Engineering Sciences*, *14*(02), 67–74.
39. Vidros, S., Kolias, C., Kambourakis, G., & Akoglu, L. (2017). Automatic detection of online recruitment frauds: Characteristics, methods, and a public dataset. *Future Internet*, *9*(1), 6.
40. Vo, M. T., Vo, A. H., Nguyen, T., Sharma, R., & Le, T. (2021). Dealing with the class imbalance problem in the detection of fake job descriptions. *Computers, Materials & Continua*, *68*(1), 521–535.

致 谢

小三号黑体，加粗，居中排列，段前、段后间距各一行。“致”与“谢”之间空2汉字字符。

××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××

小四号宋体，行间距25磅，首行缩进2汉字字符，两端对齐。

×××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××

×××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××

×××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××

×××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××

附录1 代码清单

×××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××

小四号宋体，行间距20磅，首行缩进2汉字字符。

××××××××××××××××××××××××××××××××

××××××××××××××××××××××××××××××××

××××××××××××××××××××××××××××××××

××××××××××××××××××××××××××××××××

【附录中的图、表、公式等应按附录顺序单独编序，冠以“附”字，如：附图1.1；附表2.1；式（附3.1）等，编排样式参照正文。】

附录2 数据集详细说明及分析

×××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××××

××××××××××××××××××××××××××××××××

××××××××××××××××××××××××××××××××

××××××××××××××××××××××××××××××××

××××××××××××××××××××××××××××××××

附录3 实验相关补充材料