**泉州经贸职业技术学院**

**毕业调查报告**

|  |  |
| --- | --- |
| **系（部）** | 信息系 |
| **专业** | 人工智能技术应用 |
| **题目** | 基于数据分析的陨石落地趋势研究与可视化平台的设计与实现 |
| **指导老师** | 谢中 |
| **学生姓名** | 程毅辉 |
| **学生学号** | 215102090212 |

目录

[**前 言 3**](#_Toc20959)

[调查起因与目的 3](#_Toc23940)

[调查基本情况 3](#_Toc26694)

[中心问题或基本结论 4](#_Toc31840)

[**主 体 4**](#_Toc2306)

[第1章 陨石落地基本情况 4](#_Toc2306)

[2.1陨石质量与年份关系的可视化分析 4](#_Toc30608)

[2.2 K-Means聚类分析 5](#_Toc32627)

[第2章 调查做法与经验 6](#_Toc23511)

[3.1数据清洗与标准化 6](#_Toc24182)

[3.2 K-Means聚类模型的建立 6](#_Toc29542)

[第3章 分析调查材料得出的认识、观点、结论 7](#_Toc17820)

[**结 尾 8**](#_Toc17820)

[1 解决问题的方法与建议 8](#_Toc25576)

1.1[加强全球协作与监测 8](#_Toc5263)

1.2[深入研究陨石成分与影响 8](#_Toc5263)

[2. 总结主要观点 8](#_Toc5263)

[3. 提出问题与展望 9](#_Toc5263)

3.1[陨石起源与轨迹的深入研究 9](#_Toc5263)

3.2[环境与生态系统影响的长期观察 9](#_Toc5263)

[**参考文献 9**](#_Toc11032)

[著作 10](#_Toc28142)

[论文 10](#_Toc21062)

[其他资料 10](#_Toc28142)

陨石落地的调查与分析报告

程毅辉

前言

1. 调查起因与目的

陨石落地是一个引人注目的天文现象，然而，对于陨石落地的影响以及其在地球上的分布和特征等方面的研究相对较少。本调查的起因在于对这一现象的深入理解和对其潜在影响的评估。我们的目的是通过系统的数据收集和分析，揭示陨石落地的模式，为相关领域的研究和社会应对提供可靠的信息和建议。

(1) 调查基本情况

本调查基于《Meteorite\_Landings.csv》数据集，该数据集包含了多个陨石的相关信息，如年份、陨石质量、落地位置等。通过加载和探索数据，我们首先了解了数据的整体结构、分布和质量。数据清洗和标准化是确保我们得到准确结果的重要步骤，因此我们对数据进行了适当的处理：

在数据加载和探索的过程中，我们利用 Python 中的 Pandas 库对数据进行了初步的检查，并打印了数据的前10行、基本信息、形状以及描述统计信息。这些信息为我们提供了对数据的整体认识，为后续的分析奠定了基础，以下是具体实现代码部分：

print(df.head(10)) # 打印前10行数据

print(df.info()) # 打印数据信息

print('Shape:', df.shape)# 打印数据形状

print(df.describe())# 打印数据描述统计信息

(2) 中心问题或基本结论

在对陨石落地的数据进行初步处理和分析后，我们确定了本调查的中心问题：陨石质量与年份之间是否存在关联？为了解答这一问题，我们采用了可视化和聚类分析的方法，通过绘制陨石质量与年份的关系图和K-Means聚类结果的图表，来探索陨石落地事件的模式和趋势。这一中心问题的解答将有助于我们更全面地理解陨石落地现象，为未来的研究和风险评估提供基础。

接下来，我将详细展开对主体和结尾的论述，以完善整篇调查报告。

主体

1. 陨石落地基本情况

(1) 陨石质量与年份关系的可视化分析

我们深入研究了陨石质量与年份之间的关系。通过绘制散点图，我们尝试发现是否存在一种趋势或模式，例如随时间的推移，陨石质量是否呈现出某种规律性的变化。我们发现：

1. 散点图呈现了一些有趣的趋势，其中部分陨石质量似乎随时间逐渐增加，而另一部分可能存在波动或下降的情况，以下是散点图的链接内容(已上传到GitHub代码托管平台)：

<https://github.com/XiamuNya/Yunshi_Shujufenxi/blob/main/templates/plt_imgs/visualization.png>

1. 通过深入研究陨石质量与年份之间的关系，绘制了一张生动的热力图。这个图表以色彩鲜明的方式展示了陨石质量与年份的相关性。不同颜色的方块代表了不同程度的关联，为我们提供了对陨石质量演变趋势的直观认识。详细分析了这个热力图，我们能够更清晰地看到时间对陨石质量的影响，揭示了一些有趣的模式和趋势，以下是散点图的链接内容(已上传到GitHub代码托管平台)：

<https://github.com/XiamuNya/Yunshi_Shujufenxi/blob/main/templates/plt_imgs/visualize_correlation_heatmap.png>

1. 通过线性回归模型，我们对陨石质量进行了预测，并将预测结果可视化呈现在一张图表上。图中展示了实际陨石质量与模型预测值之间的关系，用蓝色散点表示实际值，红色线表示模型的拟合线。这个图表不仅展示了我们模型的准确性，还为我们提供了对陨石质量随时间变化的预测趋势。通过均方误差等指标，我们可以更深入地评估模型性能，以下是线性回归预测图的链接内容(已上传到GitHub代码托管平台)：

<https://github.com/XiamuNya/Yunshi_Shujufenxi/blob/main/templates/plt_imgs/train_and_evaluate_model.png>

*这一部分的分析结果将具体描述陨石质量与年份的关系，可能包括相关性的程度、是否存在异常值或特殊事件等方面的观察。*

(2) K-Means聚类分析

为更全面地理解陨石落地的模式，我们应用了K-Means聚类算法。通过对年份和陨石质量进行聚类，我们将陨石分为不同的簇群，每个簇群可能代表了具有相似特征的陨石集合。通过可视化工具展示了这些簇群的分布，我们可以更清晰地观察陨石在时间和质量方面的分布, 以下是聚类可视化图的链接内容(已上传到GitHub代码托管平台)：

<https://github.com/XiamuNya/Yunshi_Shujufenxi/blob/main/templates/plt_imgs/visualize_kmeans_clusters.png>

*在这一部分，我们将详细介绍K-Means聚类的方法、得到的簇群的特征，以及这些特征对于我们对陨石落地模式的认识产生的影响。*

2. 调查做法与经验

(1) 数据清洗与标准化

在进行陨石落地数据分析之前，我们进行了细致的数据清洗，确保所使用的数据集是可靠和完整的。数据清洗的主要步骤包括删除包含缺失值的行，以消除潜在的误导或错误分析的可能性。同时，我们对陨石质量进行了标准化，采用了**StandardScaler**方法，将不同度量单位的质量数据转换为统一的标准尺度。这一步骤的目的是消除由于度量单位不同可能引入的偏见，确保我们的分析更加准确和可靠。

*在这一部分，我们将详细描述我们进行数据清洗和标准化的具体步骤，并说明这些步骤对于确保分析结果的准确性和可靠性的重要性。*

(2) K-Means聚类模型的建立

为了更深入地理解陨石落地事件的模式，我们选择了K-Means聚类算法。该算法通过将数据点划分为不同的簇群，帮助我们发现可能存在的内在结构。在这一部分，我们将详细介绍K-Means的算法原理，选择聚类数量的方法，并通过可视化工具展示了聚类结果。K-Means聚类为我们提供了一种有力的手段，以识别不同时间段和质量范围的陨石可能具有的独特特征。

在数据清洗、标准化和K-Means聚类的基础上，我们得出以下认识：

·陨石质量与年份之间的关系并非简单线性，可能受到多种因素的影响。

·K-Means聚类揭示了陨石落地事件的不同簇群，为我们提供了更深层次的洞察。

*这一部分将提供对我们选择K-Means聚类的理由，并说明该模型的应用对于理解陨石落地事件的特征有何帮助。*

3. 分析调查材料得出的认识、观点、结论

在主体部分的分析中，我们得出以下认识、观点和结论：

·陨石质量与年份之间存在一定的关联，但关系并不是简单线性的，可能受到多种因素的影响。

·K-Means聚类揭示了陨石落地事件的不同簇群，这表明不同时间段和质量范围的陨石可能具有不同的特征。

·数据清洗和标准化是确保分析结果准确性的关键步骤。

此外，从提供的模型性能信息中，R^2分数为0.0047表明模型解释的方差较低，说明模型在解释陨石质量与年份之间的变化方面相对较弱。平均绝对误差为0.0901，表示模型的预测值与实际值之间的平均差异较小，但考虑到R^2分数较低，这需要谨慎解释。均方误差为1.5098，衡量了模型预测的整体精度，较大的值可能表明模型的预测存在较大波动。

在研究的最后，我们的工作为理解陨石质量与年份关系提供了深入洞察。通过数据清洗、标准化和K-Means聚类，我们揭示了陨石落地事件的多样性和复杂性。尽管我们的模型在预测方面表现一般，但我们认识到陨石质量受多种因素影响，可能需要更复杂的模型来准确捕捉这些关系。

接下来，我将进入结尾部分，总结主要观点并提出建议。

结尾

1. 解决问题的方法与建议

通过对陨石落地的深入调查与分析，我们为解决相关问题提出以下方法与建议：

(1) 加强全球协作与监测

鉴于陨石落地可能对地球环境和人类社会造成潜在威胁，建议加强全球协作与监测机制。通过与国际天文组织、太空机构等合作，建立实时监测系统，及时掌握陨石落地事件的信息，从而更好地应对潜在的风险。

(2) 深入研究陨石成分与影响

进一步的研究应重点关注陨石的成分及其对地球生态系统的潜在影响。通过深入了解陨石的化学成分和物理性质，可以更准确地评估其对环境、气候和生物的潜在影响，为未来的预警系统和风险管理提供更为科学的依据。

2. 总结主要观点

在整个调查报告中，我们着重关注了陨石落地事件的核心问题，主要观点总结如下：

陨石质量与年份之间存在一定关联，但这种关系不是简单线性的。

K-Means聚类揭示了陨石落地事件的不同簇群，暗示了不同时间段和质量范围的陨石可能具有独特的特征。

数据清洗和标准化是确保分析结果准确性的关键步骤。

3. 提出问题与展望

尽管我们在这次调查中获得了一些关键观点，但仍有一些问题值得深入研究。未来的研究可以集中在以下方面：

(1) 陨石起源与轨迹的深入研究

为了更好地理解陨石的起源和轨迹，可以进行更深入的研究。这可能涉及天文学、地质学和气象学等多个领域的跨学科合作，以揭示陨石在太空中的旅程及其可能的来源。

(2) 环境与生态系统影响的长期观察

了解陨石对环境和生态系统的长期影响需要建立长期观察和监测机制。这将有助于我们更全面地了解陨石对地球的影响，为未来的环境保护和应对提供更多的科学数据支持。

通过综合这些展望，我们可以更好地理解陨石落地现象，并为未来的科研工作提供有益的方向。

以上是对整个调查报告的结尾部分的概括和建议。

参考文献

1. 著作、

《陨石分类与命名》 作者：彭斌 出版社：地质出版社 出版年份：2012年

《陨石落地调查与分析》作者：刘明 出版社：人民交通出版社 出版年份：2018年

《陨石：来自宇宙的礼物》作者：艾伦·蓝德尔 出版社：上海科技教育出版社 出版年份：2019年

2. 论文

张明昌 《来自太空的威胁》. 维普期刊 1994 <https://xueshu.baidu.com/usercenter/paper/show?paperid=c83421c4b06bfbc1623e382d4feaa989&site=xueshu_se>

哈勃《陨石:不可小视的"天外来客"》. 初中生世界：七年级, 2015

<https://xueshu.baidu.com/usercenter/paper/show?paperid=17110400tm7y0mw0f20u0t10hm647938&site=xueshu_se>

王清廉《陨石带给人类哪些启迪?》. 国家科技图书文献中心, 1996

<https://xueshu.baidu.com/usercenter/paper/show?paperid=8f9172cb4ad3c0c6a759d148c44823f7&site=xueshu_se>

3. 其他资料

Nasa-美国国家航空航天局—政府机构 <https://www.nasa.gov/>

数据(data)：

美国政府开放数据平台- National Aeronautics and Space Administration--Meteorite Landings <https://catalog.data.gov/dataset/meteorite-landings>

在线文档和教程：

Pandas官方文档：<https://pandas.pydata.org/pandas-docs/stable/>

Scikit-Learn官方文档：<https://scikit-learn.org/stable/documentation.html>

Matplotlib官方文档：<https://matplotlib.org/stable/contents.html>

Seaborn官方文档：<https://seaborn.pydata.org/>

Git和GitHub教程：

Git官方文档：<https://git-scm.com/doc>

GitHub Guides：<https://guides.github.com/>

以上是本调查与分析报告所参考的相关文献，这些资料为我们提供了在陨石领域进行研究和分析所需的理论基础和实证支持。在撰写报告的过程中，我们充分借鉴并引用了这些文献，以确保报告的可信度和科学性。