零样本学习在过时风险预测中的应用 ⋆

Elie Saad [*,**] **Aya Mrabah** [*] **Mariem Besbes** [*]
Marc Zolghadri [*] **Victor Czmil** [**] **Claude Baron** [***]
Vincent Bourgeois [**]

[*] *ISAE-Supméca, 3 Rue Fernand Hainaut, 93400 Saint-Ouen-sur-Seine, France (e-mail: {elie.saad,aya.mrabah,mariem.besbes,marc.zolghadri}@isae-supmeca.fr).*
[**] *SNCF - Campus Fruitier (Eurostade), 6 Av. François Mitterrand, 93210 Saint-Denis, France (e-mail: {elie.saad,victor.czmil,vincent.bourgeois1}@reseau.sncf.fr)*
[***] *LAAS-CNRS, 7 Av du Colonel Roche, 31400, Toulouse, France (e-mail: claude.baron@insa-toulouse.fr)*

摘要组件过时在依赖电子组件的行业中带来了重大挑战，导致成本增加以及系统安全性和可用性的中断。准确预测过时风险至关重要，但由于缺乏可靠数据而受到阻碍。本文提出了一种使用零样本学习（ZSL）和大型语言模型（LLMs）来利用表格数据中的特定领域知识解决数据限制的新型过时风险预测方法。该方法应用于两个实际数据集，展示了有效的风险预测能力。对四种 LLMs 进行比较评估强调了选择适合特定预测任务的正确模型的重要性。

Keywords: 过时预测, 零样本学习, 大型语言模型

1. 介绍

过时是对依赖电子元件和复杂系统的行业的一个重大挑战。当产品由于技术进步、市场变化或新法规 (International Electrotechnical Commission, 2019) 而变得过时或不可用时，就会发生这种情况，导致维护成本增加，系统 (Zolghadri et al., 2023) 的安全性和可用性中断，以及运营效率低下 (Mellal, 2020)。

主动过时管理是一种战略方法，用于解决长期生命周期系统中的组件过时问题，特别是在复杂产品系统 (Meng et al., 2014) 中。它涉及监控和优先处理整个产品结构，以提供短期和中期的解决方案，例如通过预测过时风险来实现。由于技术的迅速发展，预测过时风险变得至关重要，特别是对于长期系统 (Jennings et al., 2016) 而言。传统方法通常依赖于主观输入和繁琐的市场分析，导致预测结果不一致。为了解决这些问题，开发了基于机器学习的方法来进行更准确、更高效的过时预测 (Jennings et al., 2016)。

改进预测技术并采取主动管理措施以应对组件过时 (Rust et al., 2022) 的挑战至关重要。在过时预测中，特别是对于电子元件而言，一个重大问题是缺乏足够的可靠数据，因为公司可能不会一致记录过时问题 (Gupta et al., 2021)。这包括不完整数据集和难以获取相关信息等问题 (Zolghadri et al., 2023)。此外，可用数据的局限性进一步限制了预测工作的有效性 (Moon et al., 2022)。

零样本学习（ZSL）为解决有限标记数据的挑战提供了一种有前景的解决方案，这指的是在各个领域中每个实例都标注了特定信息或类别以用于训练机器学习模型的数据集。这是一种可以在未见过的类别上对数据进行分类的机器学习方法，减少了对可用数据 (Riordan

⋆ This work was supported by SNCF RESEAU.

et al., 2024) 的需求。ZSL 是一种专门形式的知识迁移学习 (Lazaros et al., 2024)，即利用从一个任务中获得的知识来改进与之相关的新的任务的学习。表格数据集富含领域特定信息，为简单的知识迁移学习 (Wang, 2024) 带来了挑战，需要定制的方法和结合领域知识以实现有效的预测。为了解决这一问题，最近的研究探索了使用大型语言模型（LLMs）对表格数据进行分类的零样本学习方法，这些方法利用了 LLMs 中编码的先验知识 (Mandal et al., 2024; Sui et al., 2024)。尽管取得了这些进展，但研究表明在不同任务上结果有所不同 (Hegselmann et al., 2023)。

受这些发展启发，本文探讨了如何利用近期的 LLMs 使用表格数据来预测过时风险。通过两个用例数据集评估了这些 LLMs 的有效性：一个系统数据集和一个电子元件数据集。据我们所知，之前没有工作涉及用于过时预测的 ZSL。本工作的贡献有两方面：

- 引入了一种新颖的方法，用于预测零件的状态，而无需事先具备数据。
- 进行零样本学习环境下最近发布的大型语言模型的比较评估。

本研究中使用的代码和数据集在 GitHub[1] 上公开可用，确保了可重复性并支持该领域的进一步发展。

论文的其余部分组织如下：第 2节回顾了现有的过时风险预测方法。第 3节解释了研究方法。在第 4节中，展示了实验结果并讨论了模型性能。最后，第 5节总结了论文并提出了未来研究的方向。

2. 文献回顾

过时风险预测领域越来越重视机器学习技术，特别是用于预测零件变得过时的可能性。正如 Jennings et al. (2016) 所强调的那样，监督学习通常比无监督方法更受欢迎，因为前者能够利用已知的数据状态并避免模糊的聚类结果。他们使用从 GSM Arena 网站抓取的数据集进行的研究表明，随机森林是过时风险预测中最有效的算法。这一结论得到了 Grichi et al. (2017) 的支持，他认为随机森林对数据集变异和偏差的抗性使其非常适合此类任务。在此基础上，Grichi et al. (2018a) 和 Grichi et al. (2018b) 将元启发式遗传算法和粒子群优化算法整合到随机森林框架中，提升了其预测性能。

[1] https://github.com/Inars/零样本学习在过时风险预测中的应用

进一步的支持来自 Trabelsi et al. (2021)，他们评估了多种特征选择方法（过滤器、封装器和嵌入式）在五种机器学习算法中的表现，包括人工神经网络和支持向量机。他们的研究结果呼应了早期的研究，证实随机森林在预测过时风险方面达到了最高的精度。与之前的研究一样，他们的分析也依赖于从 GSM Arena 抓取的技术数据。

一种替代的两阶段预测方法由 Liu and Zhao (2022) 提出。该方法结合了特征选择中的 ELECTRE I 方法和增强型径向基函数神经网络进行预测。改进粒子群优化、梯度下降增强和数据加权技术等创新显著提高了聚类效果、收敛速度和预测准确性。

尽管文献中讨论的现有方法在过时风险预测方面表现出显著的进步，但它们都依赖于大量的可用数据进行训练。这种对广泛数据集的依赖可能成为数据不完整、无法访问或不可用场景下的限制。本文提出的方法通过引入一种无需事先访问标记数据即可预测零件状态的新能力来解决这一问题。

3. 方法论

本节概述了用于评估 LLMs 在预测系统和组件可用性或过时情况方面的能力的方法。它描述了第 3.1节中采用的框架。其余部分介绍了第 3.2节中用于比较的模型、第 3.3节中涉及的数据集以及第 3.4节中评估性能所使用的指标。

3.1 框架

本文中的方法可以概括为：用 LLM 提示系统或电子组件的各种特征，然后询问该系统或组件是否可用或已过时。这种方法是对 Hegselmann et al. (2023) 引入的框架进行了一些修改后的版本。该方法概述如下：

(1) **序列化：** 表格数据被序列化为自然语言字符串表示。具体来说，采用“文本模板”序列化方法，每个特征以如下格式进行文本表示：“该*列名*是*值*。”
(2) **提示大语言模型：** 然后将序列化数据用于提示 LLM，同时附上对分类任务的简要描述。该提示指示 LLM 分析提供的数据并预测新数据点的类别标签，即可用或已废弃。该提示遵循以下结构：“二极管/手机特性: *序列化*。问题: 这个二极管/手机是否可用？是或否？答案:”，其中“二极管”用于 Arrow 数据集，“手机”用于 GSM Arena 数据集。

(3) **评估** 基于大规模语言模型的分类器在完整的 Arrow 和 GSM 竞技场数据集上的表现进行了评估，没有任何微调。

3.2 模型

为了识别最佳的零样本学习模型以预测系统和组件的状态，选择了四个近期知名的模型进行评估：Hugging Face 的 Task Zero (T0)、Meta 的 Llama3、Google 的 Gemma2 和 Microsoft 的 Phi3。

第一个模型，T0，由 Sanh et al. (2022) 引入。作者们考察了多任务学习在使 LLMs 能够泛化到广泛的任务范围中而不需特定任务训练的潜力，这意味着该模型不需要针对某个具体问题或应用进行微调或训练。这种方法利用了一个预训练的自动编码器模型，并对其进行了多样数据集上的微调，这些数据集被重新表述为自然语言提示。这些提示以人类可读的格式提供任务特定指令，使模型能够理解和响应各种任务，甚至包括那些它之前未曾遇到的任务。该模型的架构由编码器-解码器框架组成，其中编码器处理输入提示，解码器生成相应的输出，从而促进模型适应新任务的能力。通过在这个多任务混合数据集上进行微调，该模型实现了强大的零样本性能，通常在标准基准测试中超越了更大的模型。这种能力对于预测零件状态尤为重要，因为模型可以利用从多样化任务中学到的通用知识来理解并预测过时风险，而无需特定领域的大量标注数据集。

第二个模型，Llama3，由 Dubey et al. (2024) 引入，代表了一组新的大型语言模型（LLM），旨在处理多语言任务、编码、推理和工具使用。这些模型基于转换器架构构建，这是一种神经网络设计，通过自注意力机制来高效处理输入序列，通过对数据的不同部分分配不同的重要性水平来实现这一点。这种架构使模型能够理解序列内的关系，无论它们的位置如何。Llama3 的最大版本包含 3.21 亿个参数，并支持多达 128K 个标记的上下文窗口，其中标记代表由模型处理的最小文本单元——如词、子词或字符。这些标记是将语言分解为计算管理单位的基础，使模型能够生成连贯且语境合适的响应。Llama3 模型通过两个阶段进行训练：首先，在多样化和干净的数据集上预训练以理解语言结构并获得知识；其次，通过后训练来微调模型以遵循指令的行为。这种遵循指令的能力特别有助于预测某个部分的状态，因为它使模型能够解释与过时风险相关的提示，并利用其广泛的预训练知识生成准确的预测，即使没有特定领域的训练数据也是如此。

所使用的第三个模型是 Team et al. (2024) 介绍的 Gemma2 模型。Gemma2 模型系列引入了一系列轻量级、开放的语言模型，旨在平衡高性能与实际部署限制。这些模型参数范围从 2 亿到 27 亿，利用了诸如知识蒸馏等关键技术革新，即训练较小的模型以复制较大模型的表现，从而提高了效率和效果。该技术使 Gemma2 模型能够相对于大小为其 2 到 3 倍的模型实现具有竞争力的表现。此外，这些模型采用了架构改进，包括分组查询注意力和局部全局注意力策略，优化了信息处理。这些特性使得 Gemma2 模型特别适合于预测零件的状态，因为其高效性和先进的注意力机制可以快速分析表格数据，并识别出表明过时风险的关键模式。

最后，Phi3 模型，详细描述为 Abdin et al. (2024)，引入了 phi-3-mini，这是一个参数量高达 3.8 亿的高效语言模型，在性能上超越了诸如 Mixtral8x7B(Jiang et al., 2024) 和 GPT-3.5 等更大规模的模型，尽管它的尺寸较小。关键创新不在于增加模型的大小，而在于增强训练数据集，该数据集结合了经过严格筛选的网络数据和由语言模型生成的合成数据。该模型的架构基于一个具有 3072 个隐藏维度、32 个注意力头和 32 层的解码器变换器。它支持长度为 4K 个标记的上下文，并使用了一个词汇量大小为 $320,641$ 的分词器。此设计优先考虑计算效率和强大的泛化能力，使 Phi3 特别适合预测零件的状态。它依赖高质量的数据来识别报废风险中的细微模式，即使在计算资源受限的情况下也能提供有效的可扩展解决方案，以应对现实世界的预测挑战。

3.3 数据集

为了评估所提出的方案，分析了两个不同的案例研究。第一个涉及使用 GSM Arena 数据集进行的系统过时的高级研究，重点是智能手机的过时情况，如 Jennings et al. (2016) 所强调的。第二个案例研究是在组件级别上使用的 Arrow 数据集进行的过时研究，该数据集提供了齐纳二极管这种电子半导体元件的数据。关于齐纳二极管的信息来源于 Electronics (2024) 网站，该公司分销电子组件。

两个数据集的各种规格表示在表 1中。列 N 表示数据集中总行数。列 N_0 和 N_1 分别表示废弃实例和可用实例

的总数。反斜杠 (\) 列表示数据集中废弃实例的比例百分比。

Arrow 数据集包含 11,080 个实例，其中 7,580 个被标记为过时的，3,500 个为可用的，从而导致过时实例的比例为 68.41%。另一方面，GSM Arena 数据集包含 8,628 个实例，其中 4,773 被标记为过时的，3,855 为可用的，过时实例的比例为 55.32%。两个数据集都显示出过时和可用实例数量之间的不平衡。特别是 Arrow 数据集高度不平衡，过时实例的比例较大（68.41%），而 GSM Arena 数据集虽然不太失衡，但仍倾向于更高的过时实例比例。然而，由于本研究采用零样本学习设置，在此设置中模型在未针对这些数据集中的标记实例进行训练的情况下做出预测，因此高比例的过时案例不会影响模型的结果。这确保了推理阶段的数据不平衡不影响评估结果。

3.4 指标

为了考察四个大语言模型的表现，采用了以下五个评价指标 (Obi, 2023)：

- **准确率：** 准确率衡量的是正确分类的实例（真阳性与真阴性）占总实例的比例：

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}}, \tag{1}$$

  其中 TP（真阳性）是正确预测为正类别的实例。TN（真阴性）是正确预测为负类别的实例。FP（假阳性）是被错误分类为正类别的负类别实例。FN（假阴性）是被错误分类为负类别的正类别实例。准确率范围从 0 到 1，1 表示完全正确的分类。

- **精度：** 精确率在最小化假阳性（错误地将负实例预测为正）至关重要时应用。它确保了预测的正实例极有可能真正为正。精确率量化了所有被预测为正的实例中正确预测为正的比例：

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}}. \tag{2}$$

  精确率范围从 0 到 1，数值越高表示假阳性越少。精确率的理想值是 1，这意味着所有被预测为正的实例都被正确识别为过时组件，且没有假阳性。

表 1. 用例数据集规范

数据集	N	N_0	N_1	\
Arrow	11080	7580	3500	68.41
GSM Arena	8628	4773	3855	55.32

- **回忆：** 查全率，也称为灵敏度或真正例率（TPR），在需要最小化假阴性（未能识别正实例）的情况下使用。该指标侧重于捕获所有相关的正实例，即使这会导致一些假阳性。查全率衡量的是所有实际正实例中正确预测的正实例的比例：

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}}. \tag{3}$$

  查全率范围从 0 到 1，数值越高表示假阴性越少。查全率的理想值是 1，这表明所有过时组件都被正确识别，没有假阴性。

- **F1 分数：** 当需要在精度和召回率之间取得平衡时，会使用 F1 分数。它将两个指标合并为一个单一值，代表了正确识别阳性实例的能力与检测所有阳性实例能力之间的权衡。F1 分数是精确率和召回率的调和平均值：

$$\text{F1} = 2 \cdot \frac{\text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}} \tag{4}$$

  F1 分数范围从 0 到 1，1 表示完美的精确率和召回率。F1 分数的理想值是 1，这表明模型同时具有高精度和高召回率，假阳性和假阴性都很少。

- **曲线下的面积：** 曲线下面积（AUC）指标基于接收者操作特征（ROC）曲线评估分类器区分类别的能力，该曲线绘制了不同阈值设置下的真正例率（TPR）与假正例率（FPR）：

$$\text{FPR} = \frac{\text{FP}}{\text{FP} + \text{TN}}. \tag{5}$$

  AUC 表示 ROC 曲线下面积：

$$\text{AUC} = \int_0^1 \text{TPR}(\text{FPR}) d\text{FPR} \tag{6}$$

  AUC 的范围从 0.5（随机猜测）到 1（完美区分）。值大于 0.9 表示分类性能很强，而值在 0.7 和 0.9 之间则表明中等性能。AUC 的理想值是 1，这表示在任何阈值下都没有假阳性和假阴性的完美分类能力。

每个数据集的最佳模型是通过简单的投票技术确定的，选择在多个指标上表现最高的模型。对于 GSM Arena 数据集，选择了整体性能最优越的模型，而对于 Arrow 数据集，则根据类似的标准独立识别出最佳模型。这种方法确保为每个数据集分别选出最适合的模型。

4. 实验

近期研究表明，运行 LLMs 需要大量的计算资源，包括显著的 GPU 内存和张量卸载能力 (Isaev et al., 2023;

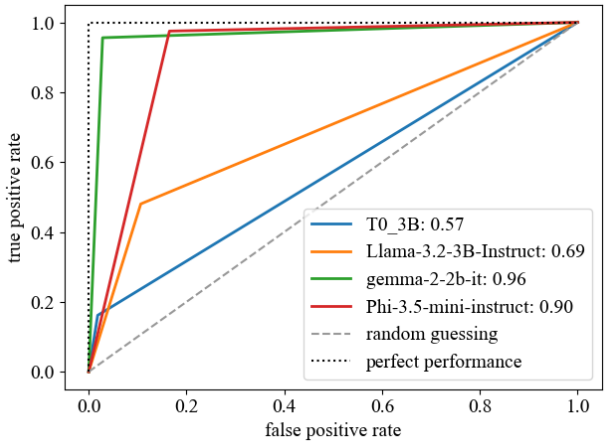

图 1. Arrow 数据集的接收者操作特性曲线

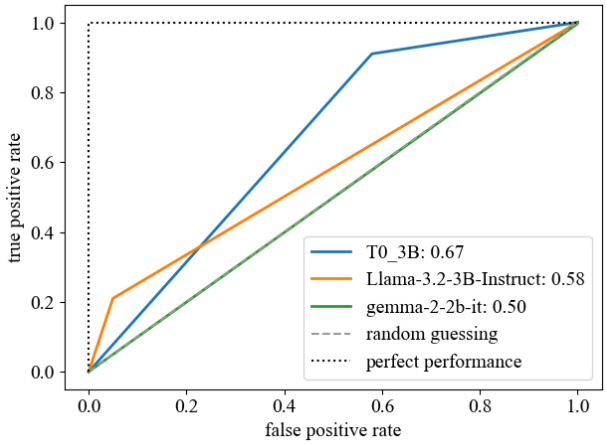

图 2. 接收者操作特征曲线用于 GSM Arena 数据集

Borzunov et al., 2024)。张量是用于在深度学习模型中表示数据的多维数组（向量）。它将矩阵（2 维数组）推广到更高维度。尽管大型模型需要大量的计算资源，但次 10-亿参数的 LLMs 提供了一种实用的替代方案，有效地平衡了计算效率和准确性，使其适用于较小的设备 (Li et al., 2024)。为了解决可访问性限制，本研究利用所选模型的小型指令调整变体（针对遵循指令行为进行微调的模型），在有可用的情况下使用。具体来说，选择的模型包括 T0_3B（2.85 亿参数）、Llama-3.2-3B-Instruct（3.21 亿参数）、Gemma-2-2B-It（2.61 亿参数）和 Phi-3.5-Mini-Instruct（3.82 亿参数）。

在实验过程中，模型 T0、Llama3.2 和 Gemma2 成功遵循了提供的指令并生成了预测结果，没有出现问题。相比之下，Phi3.5 一直未能跟随提示中的指令，即使修改了提示并尝试了各种指令结构也是如此。因此，使用第 3节提出的提示结构，Phi3.5 无法为 Arrow 数据集中的所有提示提供预测结果，错过了 276 个预测结果中的 11,080 个，并且对于 GSM Arena 数据集完全失败，没有做出任何预测（0 个预测结果中的 8,628）。这些失败的主要原因似乎是模型无法识别提示中足够的信息来做出自信的预测。

表 2中的实验结果展示了四个选定的大语言模型应用于两个用例数据集的表现。图 1和 2分别显示了大语言模型在 Arrow 数据集和 GSM Arena 数据集上的 ROC 曲线。在这项研究中，TPs 指的是模型正确预测某个部件可用的情况，而 TNs 指的是模型正确预测某个组件已过时的情况。

对于 Arrow 数据集，Gemma2 在多个指标上优于所有其他模型，实现了 96.67% 的准确率、0.94 的精度和 0.956 的召回率。它还以高 AUC 值 0.964 领先，表明其强大的分类能力。T0 模型虽然表现出不错的性能，准确率为 72.26%，但在召回率（0.161）方面表现不佳，表明其预测中假阴性率较高。Llama3.2 模型提供了更好的精度（0.676），但召回率较低（0.48），表明倾向于漏掉许多阳性实例。Phi3.5 模型在 Arrow 数据集上的性能

表 2. 实验结果

模型	度量	箭头	GSM 领土
T0	Accuracy	72.26	**69.14**
	Precision	0.803	0.66
	Recall	0.161	0.91
	F1	0.269	**0.765**
	AUC	0.572	**0.665**
llama 3.2	Accuracy	76.3	54.08
	Precision	0.676	**0.84**
	Recall	0.48	0.21
	F1	0.561	0.336
	AUC	0.687	0.58
盖玛 2	Accuracy	**96.67**	55.12
	Precision	**0.94**	0.552
	Recall	**0.956**	**0.996**
	F1	**0.948**	0.711
	AUC	**0.964**	0.498
Phi 3.5	Accuracy	72.26	—
	Precision	0.803	—
	Recall	0.161	—
	F1	0.269	—
	AUC	0.572	—

与 T0 相同，显示出相同的精确率和召回率，但 AUC (0.572) 显著较低，这表明其区分类别的能力有限。

对于 GSM Arena 数据集，T0 在准确性（69.14%）和 F1 分数（0.765）方面领先，表明它在精确度和召回率之间取得了良好的平衡。然而，Llama3.2 在精确度（0.84）方面表现出色但召回率较低（0.21），这意味着它能够正确识别许多阳性结果但也遗漏了其中相当一部分。Gemma2，在召回率方面表现良好 (0.996)，但未达到具有竞争力的准确率或 AUC 值，这反映了灵敏度和整体分类能力之间可能存在的权衡。

使用第 3节中概述的投票技术，基于多个指标的表现来识别每个数据集的最佳模型，进而识别出每项任务的最佳模型。对于 Arrow 数据集，Gemma 2 表现出色，使其成为预测电子组件状态的最佳选择。相反，在 GSM Arena 数据集中，T0 模型被认定为最佳表现模型，因此在系统状态预测任务中更受欢迎，尽管其精度不是特别显著。

5. 结论

本文提出了一种利用 ZSL 和大语言模型预测组件过时风险的新方法。通过利用表格数据集中的特定领域知识，本文展示了大语言模型可以有效应对过时预测中标签数据有限的挑战。通过对两个实际数据集的应用，一个涉及系统，另一个专注于电子组件，结果突显了在无需大量收集数据或对标注实例进行预先训练的情况下，ZSL 增强过时风险预测准确性的潜力。

在零样本学习（ZSL）背景下，四种近期大语言模型的比较评估显示了不同的性能水平，突显了为特定过时预测任务选择合适模型的重要性。

值得注意的是，Google 的 Gemma2 模型在 Arrow 数据集上优于其同类模型，而 Hugging Face 的 T0 模型则在 GSM Arena 数据集上表现出色。

未来在此领域的研究应探讨在少量样本学习框架内，对这些模型进行微调以预测系统和电子元件过时的潜力。此外，探索将大语言模型与传统的过时预测方法相结合的混合方法，可能会进一步提高过时风险预测的准确性和稳健性。

参考文献

Abdin, M., Aneja, J., Awadalla, H., Awadallah, A., Awan, A.A., Bach, N., Bahree, A., Bakhtiari, A., Bao, J., Behl, H., et al. (2024). Phi-3 technical report: A highly capable language model locally on your phone. *arXiv preprint arXiv:2404.14219*.

Borzunov, A., Ryabinin, M., Chumachenko, A., Baranchuk, D., Dettmers, T., Belkada, Y., Samygin, P., and Raffel, C.A. (2024). Distributed inference and fine-tuning of large language models over the internet. *Advances in Neural Information Processing Systems*, 36.

Dubey, A., Jauhri, A., Pandey, A., Kadian, A., Al-Dahle, A., Letman, A., Mathur, A., Schelten, A., Yang, A., Fan, A., et al. (2024). The llama 3 herd of models. *arXiv preprint arXiv:2407.21783*.

Electronics, A. (2024). Arrow. URL `http://www.arrow.com/`. Accessed on June 7, 2024.

Grichi, Y., Beauregard, Y., and Dao, T. (2017). A random forest method for obsolescence forecasting. In *2017 IEEE IEEM*, 1602–1606. IEEE.

Grichi, Y., Beauregard, Y., and Dao, T.M. (2018a). Optimization of obsolescence forecasting using new hybrid approach based on the rf method and the meta-heuristic genetic algorithm. *American Journal of Management*, 18(2).

Grichi, Y., Dao, T.M., and Beauregard, Y. (2018b). A new approach for optimal obsolescence forecasting based on the random forest (rf) technique and meta-heuristic particle swarm optimization (pso). In *Proceedings of the International Conference on Industrial Engineering and Operations Management*, 26–27. Paris, France.

Gupta, A., Kumar, M., and Singh, S. (2021). Challenges in data-driven obsolescence management: A review. *Computers in Industry*, 130, 103456.

Hegselmann, S., Buendia, A., Lang, H., Agrawal, M., Jiang, X., and Sontag, D. (2023). Tabllm: Few-shot classification of tabular data with large language models. In *AISTATS*, 5549–5581. PMLR.

International Electrotechnical Commission (2019). *IEC 62402:2019 Obsolescence Management—Application Guide*. IEC.

Isaev, M., McDonald, N., and Vuduc, R. (2023). Scaling infrastructure to support multi-trillion parameter llm training. In *Architecture and System Support for Transformer Models (ASSYST@ ISCA 2023)*.

Jennings, C., Wu, D., and Terpenny, J. (2016). Forecasting obsolescence risk and product life cycle with machine learning. *IEEE Transactions on Components, Packaging and Manufacturing Technology*, 6(9), 1428–1439.

Jiang, A.Q., Sablayrolles, A., Roux, A., Mensch, A., Savary, B., Bamford, C., Chaplot, D.S., Casas, D.d.l., Hanna, E.B., Bressand, F., et al. (2024). Mixtral of experts. *arXiv preprint arXiv:2401.04088*.

Lazaros, K., Koumadorakis, D.E., Vrahatis, A.G., and Kotsiantis, S. (2024). A comprehensive review on zero-shot-learning techniques. *IDT*, 18(2), 1001–1028.

Li, X., Lu, Z., Cai, D., Ma, X., and Xu, M. (2024). Large language models on mobile devices: Measurements, analysis, and insights. In *Proceedings of the Workshop on Edge and Mobile Foundation Models*, 1–6.

Liu, Y. and Zhao, M. (2022). An obsolescence forecasting method based on improved radial basis function neural network. *Ain Shams Engineering Journal*, 13(6), 101775.

Mandal, B., Amariucai, G., and Wei, S. (2024). Initial exploration of zero-shot privacy utility tradeoffs in tabular data using gpt-4. *arXiv preprint arXiv:2404.05047*.

Mellal, M.A. (2020). Obsolescence – a review of the literature. *Technology in Society*, 63, 101347.

Meng, X., Thörnberg, B., and Olsson, L. (2014). Strategic proactive obsolescence management model. *IEEE Transactions on Components, Packaging and Manufacturing Technology*, 4(6), 1099–1108.

Moon, K.S., Lee, H.W., Kim, H.J., Kim, H., Kang, J., and Paik, W.C. (2022). Forecasting obsolescence of components by using a clustering-based hybrid machine-learning algorithm. *Sensors*, 22(9), 3244.

Obi, J.C. (2023). A comparative study of several classification metrics and their performances on data. *World Journal of Advanced Engineering Technology and Sciences*, 8(1), 308–314.

Riordan, B., Bonela, A.A., He, Z., Nibali, A., Anderson-Luxford, D., and Kuntsche, E. (2024). How to apply zero-shot learning to text data in substance use research: an overview and tutorial with media data. *Addiction*, 119(5), 951–959.

Rust, R.M., Elshennawy, A., and Rabelo, L. (2022). A literature review on mitigation strategies for electrical component obsolescence in military-based systems. *South African Journal of Industrial Engineering*, 33(1), 25–38.

Sanh, V., Webson, A., Raffel, C., Bach, S., Sutawika, L., Alyafeai, Z., Chaffin, A., Stiegler, A., Raja, A., Dey, M., et al. (2022). Multitask prompted training enables zero-shot task generalization. In *ICLR*.

Sui, Y., Zhou, M., Zhou, M., Han, S., and Zhang, D. (2024). Table meets llm: Can large language models understand structured table data? a benchmark and empirical study. In *Proceedings of the 17th ACM WSDM 2024*, 645–654.

Team, G., Riviere, M., Pathak, S., Sessa, P.G., Hardin, C., Bhupatiraju, S., Hussenot, L., Mesnard, T., Shahriari, B., Ramé, A., et al. (2024). Gemma 2: Improving open language models at a practical size. *arXiv preprint arXiv:2408.00118*.

Trabelsi, I., Zeddini, B., Zolghadri, M., Barkallah, M., and Haddar, M. (2021). Obsolescence prediction based on joint feature selection and machine learning techniques. *ICAART (2)*, 787–794.

Wang, A.X. (2024). *Data-Centric AI: Tabular Data Synthesis with Deep Generative Models*. Ph.D. thesis, Open Access Te Herenga Waka-Victoria University of Wellington.

Zolghadri, M., Besbes, M., Bourgeois, V., and Saad, E. (2023). Micro-electronic chips shortages and obsolescence: an empirical study. *Procedia CIRP*, 120, 1570–1575.