标题：使用机器学习基于气候因素预测新西兰乳制品产量

摘要

农业，尤其是奶牛养殖，是新西兰经济、文化和环境的基石。该行业未来的可持续性越来越取决于对乳制品生产与气候条件之间复杂且不断变化的关系的理解和适应。这项研究使用机器学习（ML）技术来预测乳制品产量，以应对一系列气候变量，包括降雨量、温度、湿度和日照时间。主要目标是建立一个潜在的预测模型，支持农民和行业利益相关者加强决策过程，提高乳制品企业应对不断升级的气候变化挑战的能力。这项研究利用了广泛的历史数据集，包括几十年的气候和乳制品生产记录。评估各种机器学习算法的预测准确性和计算效率，包括时间序列预测模型和回归分析。特别注意使用特征重要性排序等方法来选择对乳制品生产率有重大影响的特征。由此产生的模型有望产生有价值的见解，并促进乳制品管理实践的优化。通过本研究，我们希望为农业预测分析提供一个模板，以平衡运营效率和环境管理，并为政策和田间农业战略提供信息。

1简介

1.1新西兰乳制品

新西兰是领先的乳制品生产国，对全球乳制品市场做出了重大贡献。新西兰的乳制品行业非常强大，每年生产约210亿升牛奶，约占全球牛奶产量的3%。这个数量相当于每天为9000万人提供两份半乳制品。新西兰人口只有500万，但它是世界第八大牛奶生产国，其绝大多数乳制品（超过95%）出口到全球130多个国家。

乳制品行业在新西兰经济中发挥着至关重要的作用，并对该国的GDP做出了重大贡献。在截至2023年3月的一年中，乳制品相关活动为经济增加了113亿美元，占该国国内生产总值的3.2%。具体而言，乳制品行业贡献了80亿美元，占GDP的2.2%，而乳制品加工贡献了34亿美元，约占GDP的0.9%，超过了其他商品生产部门。2.

新西兰乳制品行业的奶牛测试和人工授精在2022/23生产季节创下历史新高，测试奶牛数量增加了2.8%，人工授精奶牛增加了381万头。乳制品公司加工了207亿升牛奶，包括18.7亿公斤牛奶固体，与上一季相比产量略有下降，但固体含量略有增加。尽管奶牛总数减少了3.46%，但全国平均库存规模从十多年前的39头增加到441头[表1.1]。

1.2气候对乳制品生产的影响

降雨、温度、湿度和日照时间等气候因素会显著影响牧场的生长和饲料质量，并最终影响牧场的牛奶产量和奶牛养殖系统的质量[4]，[5]。这些气候变量的变化可能导致产量的大幅波动，对乳制品行业的可持续性和盈利能力构成挑战[6]。

鉴于气候因素的波动性，有必要整合适应性管理战略，从准确的预测模型中提供信息。这将使奶农能够预测和应对牧场生长和质量模式的变化，保持稳定的牛奶产量和质量。

2.研究现状

2.1传统农业预测方法

传统的农业预测方法，如线性回归和时间序列分析，已被广泛用于根据历史天气数据和农场管理实践预测牛奶产量[5]。然而，这些方法往往忽视了气候因素与乳制品生产结果之间的非线性关系和相互作用[7]。尽管传统的预测方法是基本的，但它们在处理复杂的多维数据时有局限性。为了解决这个问题，我们的研究包括一种新的方法，将这些传统模型与机器学习技术相结合，以更详细地了解数据中的潜在模式。

2.2农业机器学习的进展

机器学习（ML）的最新进展，特别是在预测建模和数据分析方面，为提高农业预测的准确性和可靠性提供了新的方法[8]。机器学习算法，如人工神经网络（ANN）、支持向量机（SVM）和随机森林，可以在大数据集中捕捉复杂的模式和非线性关系，从而增强对不同气候条件下乳制品生产的预测[9]。

这项研究承认，机器学习在农业中的应用正在蓬勃发展。通过探索一系列复杂的机器学习模型，我们弥合了气候变量的动态与乳制品生产的有形指标之间的差距。

3模型和方法

3.1数据收集和预处理

这项研究将从气象机构、奶制品合作社和研究机构[9]、[10]等知名来源收集气候变量（如温度、降水、太阳辐射）和奶制品生产指标（如牛奶产量、脂肪含量、蛋白质含量）的历史数据。数据将经过严格的预处理步骤，包括数据清理、标准化和特征工程，以确保其准确性和模型训练的适用性[11]。

数据完整性对于我们的机器学习模型的成功至关重要。因此，我们的预处理还将包括估计缺失数据的步骤，确保所有数据集中气候和生产变量的全面覆盖。

3.2型号选择

我们将对多种机器学习算法进行综合评估，以确定最适合预测乳制品生产结果的模型。候选模型可能包括支持向量回归（SVR）、梯度增强器（GBM）和长短期记忆（LSTM）网络[12]。模型选择标准将优先考虑预测准确性、计算效率和可解释性，以确定乳制品实际应用的稳健预测框架[13]。

为了确保仔细的选择过程，我们的研究将通过性能指标矩阵进行比较分析。这不仅提高了准确性，还提高了每种算法的计算效率和可解释性。在研究的后期阶段，构思一个用户友好的界面，供农民与模型互动。

3.3验证和测试

所选机器学习模型的性能将通过结合训练验证测试数据集和交叉验证技术进行严格评估[14]。将数据集随机划分为训练（70%）、验证（15%）和测试（15%）子集，以评估模型在不同气候区域和生产系统中的普遍性和稳健性[15]。此外，k倍交叉验证将用于估计模型的预测不确定性，并确定偏差或过拟合的潜在来源[16]。

在验证和测试阶段，我们将模拟各种气候条件（包括极端天气事件）进行情景测试，以评估模型的稳健性。这将大大有助于其在实际场景中的适用性。

4.研究价值

4.1对奶牛养殖的影响

开发准确可靠的基于机器学习的奶牛产量预测模型对奶牛养殖的可持续性和盈利能力具有重要意义[17]。通过整合实时天气数据和农场管理信息，农民可以主动管理气候相关风险，优化资源配置，提高整体生产力和盈利能力[18]。

该预测模型整合实时气候数据的能力突显了其在帮助农民采用精准农业技术，从而优化他们应对气候相关风险方面的有效性。

4.2对农业研究的贡献

这项研究通过证明机器学习技术在预测复杂生物系统对环境变化的反应方面的有效性，为更广泛的农业科学领域做出了贡献[19]。本研究中制定的方法框架可以调整并扩展到其他农业部门，促进循证决策和可持续资源管理[20]。

我们的研究通过绘制环境因素与农业产出之间的相互作用，为数据驱动的决策提供了蓝图。这可以指导可持续做法和激励措施的发展，鼓励在农业中采用预测分析。

5结论

总之，这项研究概述了机器学习的变革潜力，以提高新西兰乳制品行业抵御气候变化破坏稳定力量的能力。通过仔细分析历史数据并应用先进的机器学习算法，本研究提出了一个预测模型，展示了农业知识与尖端技术的和谐融合。这种模式不仅有望改善奶农的决策过程，而且预示着气候智能农业的新时代。所开发模式的内在适应性为其在不同农业部门的应用铺平了道路，并增强了整个粮食系统的复原力。最终，这项研究超越了学术讨论，为可持续乳制品生产提供了切实可行的解决方案，与环境管理和经济可行性的紧迫性产生了共鸣。