题目：利用机器学习根据气候因素预测新西兰的乳制品生产

抽象

农业，特别是奶牛养殖，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错综复杂和不断变化的关系的理解和适应。本研究采用机器学习 （ML） 技术来预测乳制品产量，以响应一系列气候变量，包括降雨量、温度、湿度、日照持续时间等。主要目标是构建一个潜在的预测模型，以支持农民和行业利益相关者在气候变化不断升级的挑战中加强决策过程并增强乳品运营的弹性。 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。评估了多种机器学习算法（包括时间序列预测模型和回归分析）的预测准确性和计算效率。特别注意使用特征重要性排序等方法选择对乳品生产力有显著影响的特征。由此产生的模型有望产生有价值的见解，促进乳品管理实践的优化。通过这项研究，我们期望为农业预测分析提供一个模板，以平衡运营效率与环境管理，从而为政策和实地农业战略提供信息。

1. 引言

1.1 新西兰的乳制品

新西兰是首屈一指的乳制品生产国，为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9000万人提供两份半的乳制品。新西兰人口只有500万，是全球第八大牛奶生产国，其绝大多数乳制品（超过95%）出口到全球130多个国家[1]。

乳制品行业在新西兰经济中发挥着举足轻重的作用，对国家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 GDP 总额的 3.2%。具体而言，奶牛养殖业注入了 80 亿美元，占 GDP 的 2.2%，而乳制品加工占 34 亿美元，占 GDP 的 0.9%，超过了其他商品生产部门。[2].

在 2022/23 产季，新西兰乳制品行业的牛群检测和人工授精创下历史新高，测试奶牛数量增加了 2.8%，人工授精奶牛增加了 381 万头。乳品公司加工了207亿升牛奶，其中含有18.7亿公斤乳固体，与上一季相比，产量略有下降，但固体含量略有增加[3]。尽管奶牛总数减少了3.46%，但全国平均牛群规模上升到441头，比十年前增加了39头[表1.1]。

1.2 气候对乳制品生产的影响

降雨量、温度、湿度和日照时间等气候因素显着影响牧场生长、牧草质量，并最终影响牧场奶牛养殖系统的牛奶产量和质量[4]，[5]。这些气候变量的变化会导致产量的显着波动，对乳制品行业的可持续性和盈利能力构成挑战[6]。

考虑到气候因素的波动性，必须整合由准确预测模型提供的适应性管理战略。这将使奶农能够预测和应对牧草生长和牧草质量的变化模式，从而保持一致的牛奶产量和质量。

2. 研究现状

2.1 农业预测的传统方法

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]。 传统的预测方法虽然是基础，但在处理复杂的多维数据方面存在局限性。为了解决这个问题，我们的研究包括一种新方法，将这些传统模型与机器学习技术叠加在一起，从而对数据的基本模式提供更细致的理解。

2.2 农业机器学习的进步

机器学习（ML）的最新进展，特别是在预测建模和数据分析领域，为提高农业预测的准确性和可靠性提供了新的途径[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，能够捕获大型数据集中的复杂模式和非线性关系，从而增强了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性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

本研究将从气象机构、乳品合作社和研究机构等知名来源收集有关气候变量（例如温度、降水、太阳辐射）和乳制品生产指标（例如牛奶产量、脂肪含量、蛋白质含量）的历史数据[9]，[10]。数据将经过严格的预处理步骤，包括数据清理、归一化和特征工程，以确保其准确性和模型训练的适用性[11]。

数据的完整性对于我们的 ML 模型的成功至关重要。因此，我们的预处理还将包括一个插补缺失数据的步骤，确保全面覆盖所有数据集中的气候和生产变量。

3.2 型号选择

将对多种ML算法进行全面评估，以确定最合适的模型来预测乳制品生产结果。候选模型可能包括支持向量回归（SVR）、梯度提升机（GBM）和长短期记忆（LSTM）网络等[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，以确定一个强大的预测框架，用于乳制品的实际应用[13]。

为了确保细致的选择过程，我们的研究将包括通过性能指标矩阵进行比较分析。这不仅会提高准确性，还会提高每个算法的计算效率和易解释性。在研究的后期阶段，将概念化一个供农民与模型交互的用户友好界面。

3.3 验证和测试

所选ML模型的性能将使用训练-验证-测试数据集和交叉验证技术的组合进行严格评估[14]。数据集将被随机分为训练 （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通用性和鲁棒性[15]。此外，将采用 k 倍交叉验证来估计模型的预测不确定性并识别偏差或过拟合的潜在来源[16]。

在我们的验证和测试阶段，我们将实施情景测试，模拟各种气候条件，包括极端天气事件，以评估我们模型的鲁棒性。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

开发一种准确可靠的基于ML的乳制品生产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，农民可以主动管理与气候相关的风险，优化资源配置，并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使农民能够采用精准农业技术方面的效用，从而优化了他们对气候相关风险的反应。

4.2 对农业研究的贡献

这项研究通过证明ML技术在预测复杂生物系统对环境变异性的响应方面的功效，为更广泛的农业科学领域做出了贡献[19]。本研究中开发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]。

通过绘制环境因素与农业产出之间的相互作用，我们的研究为数据驱动的政策制定提供了蓝图。这可以指导制定可持续的做法和激励措施，鼓励在农业中采用预测分析.

5. 结论

总之，这项研究概括了机器学习在加强新西兰乳制品行业抵御气候变化不稳定力量方面的变革潜力。通过细致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，证明了农业知识与尖端技术的和谐融合。该模式不仅有望提升奶农的决策过程，还预示着气候智能型农业的新时代。所开发模型的内在适应性为其在不同农业领域的应用铺平了道路，增强了整个粮食系统的复原力。最终，这项研究超越了学术话语，为可持续乳制品生产提供了务实的解决方案，与环境管理和经济可行性的紧迫性产生共鸣。

Title: Using machine learning to predict dairy production in New Zealand based on climate factors

abstract

Agriculture, particularly dairy farming, is a cornerstone of New Zealand's economy, culture and environment. The future sustainability of the industry increasingly depends on understanding and adapting to the complex and changing relationship between dairy production and climate conditions. This study employs machine learning (ML) techniques to predict dairy yields in response to a range of climate variables, including rainfall, temperature, humidity, sunshine duration, etc. The main objective is to build a potential predictive model to support farmers and industry stakeholders in enhancing decision-making processes and increasing the resilience of dairy operations amid the escalating challenges of climate change. The study drew on an extensive historical data set, including decades of climate and dairy production records. The predictive accuracy and computational efficiency of a variety of machine learning algorithms, including time series forecasting models and regression analysis, are evaluated. Special attention is paid to selecting features that have a significant impact on dairy productivity using methods such as feature importance ranking. The resulting model is expected to generate valuable insights and facilitate the optimization of dairy management practices. Through this research, we expect to provide a template for agricultural predictive analytics to balance operational efficiency with environmental stewardship to inform policy and on-the-ground agricultural strategies.

1 Introduction

1.1 New Zealand dairy products

New Zealand is a leading dairy producer and contributes significantly to the global dairy market. New Zealand's dairy industry is very strong, producing approximately 21 billion liters of milk annually, accounting for approximately 3% of global milk production. This amount is equivalent to providing 90 million people with two and a half servings of dairy products every day. With a population of only 5 million, New Zealand is the eighth largest milk producer in the world, and the vast majority of its dairy products (more than 95%) are exported to more than 130 countries around the world [1].

The dairy industry plays a pivotal role in New Zealand's economy and makes a significant contribution to the country's GDP. Dairy-related activities added $11.3 billion to the economy in the year to March 2023, accounting for 3.2% of the country's total GDP. Specifically, dairy farming injected $8 billion, or 2.2% of GDP, while dairy processing accounted for $3.4 billion, or 0.9% of GDP, outpacing other commodity-producing sectors. [2].

The New Zealand dairy industry's herd testing and artificial insemination hit record highs during the 2022/23 season, with a 2.8% increase in the number of cows tested and 3.81 million more artificially inseminated cows. Dairy companies processed 20.7 billion liters of milk, which contained 1.87 billion kilograms of milk solids, a slight decrease in production compared to the previous season, but a slight increase in solids content [3]. Although the total number of dairy cows decreased by 3.46%, the national average herd size increased to 441 cows, 39 cows more than a decade ago [Table 1.1].

1.2 Impact of climate on dairy production

Climatic factors such as rainfall, temperature, humidity and daylight hours significantly affect pasture growth, forage quality, and ultimately milk production and quality in pasture dairy farming systems [4], [5]. Changes in these climate variables can lead to significant fluctuations in yields, posing challenges to the sustainability and profitability of the dairy industry [6].

Given the volatility of climatic factors, it is essential to integrate adaptive management strategies informed by accurate predictive models. This will enable dairy farmers to predict and respond to changing patterns of pasture growth and pasture quality, maintaining consistent milk production and quality.

2. Research status

2.1 Traditional methods of agricultural forecasting

Traditional agricultural forecasting methods, such as linear regression and time series analysis, have been widely used to predict milk production based on historical weather data and farm management practices [5]. However, these methods often ignore the non-linear relationships and interactions between climatic factors and dairy production outcomes [7]. Although traditional forecasting methods are fundamental, they have limitations in handling complex multidimensional data. To address this problem, our research includes a new approach that overlays these traditional models with machine learning techniques to provide a more nuanced understanding of the underlying patterns in the data.

2.2 Progress in agricultural machine learning

Recent advances in machine learning (ML), especially in the fields of predictive modeling and data analysis, provide new ways to improve the accuracy and reliability of agricultural forecasts [8]. Machine learning algorithms, such as artificial neural networks (ANN), support vector machines (SVM), and random forests, are able to capture complex patterns and nonlinear relationships in large data sets, thereby enhancing predictions of dairy production under different climate conditions [9 ].

This study acknowledges that the application of machine learning in agriculture is booming. By exploring a series of sophisticated machine learning models, we bridge the gap between the dynamics of climate variables and tangible indicators of dairy production.

3. Models and methods

3.1 Data collection and preprocessing

This study will collect historical data on climate variables (e.g. temperature, precipitation, solar radiation) and dairy production indicators (e.g. milk yield, fat content, protein content) from reputable sources such as meteorological agencies, dairy cooperatives and research institutions [9] ,[10]. The data will undergo rigorous pre-processing steps, including data cleaning, normalization and feature engineering, to ensure its accuracy and suitability for model training [11].

Data integrity is critical to the success of our ML models. Therefore, our preprocessing will also include a step to impute missing data, ensuring comprehensive coverage of climate and production variables in all datasets.

3.2 Model selection

A comprehensive evaluation of multiple ML algorithms will be conducted to determine the most appropriate model to predict dairy production outcomes. Candidate models may include support vector regression (SVR), gradient boosting machine (GBM), and long short-term memory (LSTM) networks [12]. Model selection criteria will prioritize prediction accuracy, computational efficiency, and interpretability to identify a robust prediction framework for practical applications in dairy products [13].

To ensure a meticulous selection process, our research will include a comparative analysis through a matrix of performance indicators. This will not only improve accuracy, but also increase the computational efficiency and interpretability of each algorithm. In later stages of the research, a user-friendly interface for farmers to interact with the model will be conceptualized.

3.3 Verification and testing

The performance of the selected ML models will be rigorously evaluated using a combination of training-validation-test datasets and cross-validation techniques [14]. The data set will be randomly divided into training (70%), validation (15%) and test (15%) subsets to evaluate the generalizability and robustness of the model in different climate regions and production systems [15]. Additionally, k-fold cross-validation will be employed to estimate the predictive uncertainty of the model and identify potential sources of bias or overfitting [16].

During our validation and testing phase, we will implement scenario tests simulating a variety of climate conditions, including extreme weather events, to assess the robustness of our model. This will greatly contribute to its applicability in practical scenarios.

4. Research value

4.1 Impact on dairy farming

Developing an accurate and reliable ML-based dairy production prediction model is of great significance to the sustainability and profitability of dairy farming [17]. By integrating real-time weather data and farm management information, farmers can proactively manage climate-related risks, optimize resource allocation, and improve overall productivity and profitability [18].

The predictive model’s ability to integrate real-time climate data highlights its utility in enabling farmers to adopt precision agriculture technologies, thereby optimizing their responses to climate-related risks.

4.2 Contribution to agricultural research

This study contributes to the broader field of agricultural science by demonstrating the efficacy of ML techniques in predicting the response of complex biological systems to environmental variability [19]. The methodological framework developed in this study can be adapted and extended to other agricultural sectors, promoting evidence-based decision-making and sustainable resource management [20].

By mapping the interactions between environmental factors and agricultural output, our research provides a blueprint for data-driven policy development. This can guide the development of sustainable practices and incentives that encourage the adoption of predictive analytics in agriculture.

5 Conclusion

In summary, this study outlines the transformative potential of machine learning to strengthen the New Zealand dairy industry’s resilience to the destabilizing forces of climate change. By meticulously analyzing historical data and applying advanced machine learning algorithms, this study proposes a predictive model that demonstrates the harmonious integration of agricultural knowledge and cutting-edge technology. Not only does the model promise to improve dairy farmers’ decision-making processes, it also heralds a new era of climate-smart farming. The inherent adaptability of the developed model paves the way for its application in different agricultural sectors, enhancing the resilience of the entire food system. Ultimately, this research transcends academic discourse to provide pragmatic solutions for sustainable dairy production that resonate with the urgency of environmental stewardship and economic viability.