pHase

Fermented Raw Milk Stage Prediction Model

CSI 5170/4170 - Machine Learning
December 8th, 2025
By: Nate Oberdier

Introduction & Project Goals

Problem Definition

- Raw milk ferments naturally through predictable stages
- No ML model exists to estimate fermentation stage or pH
- Existing research focuses on pasteurized, lab-controlled dairy
- Consumers rely on guesswork (smell, texture)

Relevance to Machine Learning

- Nonlinear time-series modeling
- Regression + classification
- Real-world biological data
- Feature engineering & evaluation

Goals & Scope

- Predict pH from time, temperature, and culture conditions
- Predict discrete fermentation stages
- Build a real-time interactive prediction tool
- Scope: 12-jar dataset, 45 days, 67-69°F environment

Background & Literature

Existing Research Themes

- Yogurt/kefir studies: pasteurized milk, standard cultures
- Industrial fermentation is highly controlled
- Spoilage detection: fresh vs spoiled classification

Key Gaps

- Lack of raw milk datasets at household temperatures
- No ML models for full fermentation stages
- Rarely models natural microbial diversity
- Consumer prediction tools do not exist

pHase Contribution

- Real raw milk dataset
- Full stage modeling (fresh to cheese)
- Focus on natural fermentation, not industrial

Methodology Pt. 1

Study Design

- 12 jars of fresh raw milk
- Jars 1–6: no culture
- Jars 7–12: 1 tbsp naturally fermented milk added
- Stored 45 days at 67–69°F

Data Collection

- 50 measurements per jar -> 600 total entries
- pH measured together for consistency
- High-frequency sampling early in fermentation
- Slower sampling later as changes slowed

Recorded Variables

- Date, time, jar ID, pH, culture status

Methodology Pt. 2

Preprocessing

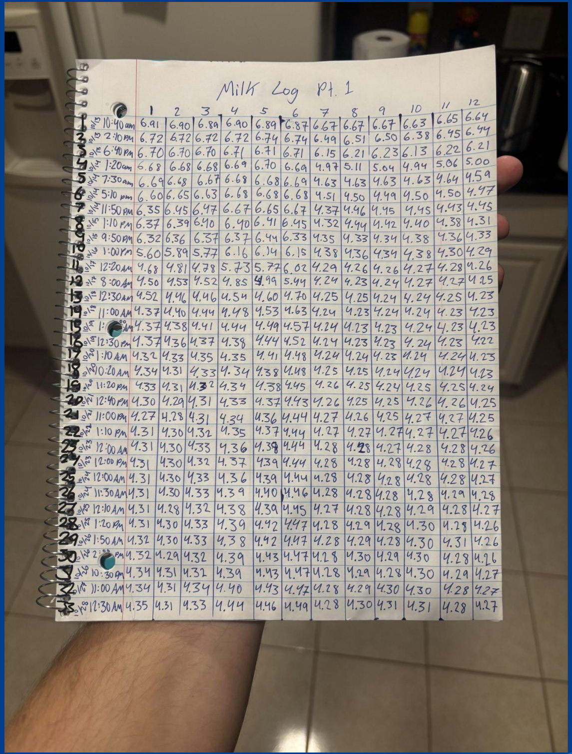
- Merge date + time -> timestamp
- Convert to hours since fermentation start
- Exponential smoothing to reduce noise
- Engineered features:
  - Rate of pH change, Time², pH × time

Stage Labeling

- Based on known pH ranges + biological order
- Stages forced to move forward only

Models Used

- Random Forest Regressor (pH)
- Random Forest Classifier (stage)
- Train/val/test split by jar and culture group (4/1/1)



Evaluation: Research Question

Main Research Question:

- Can raw milk fermentation be predicted accurately using ML?

Specific Focus

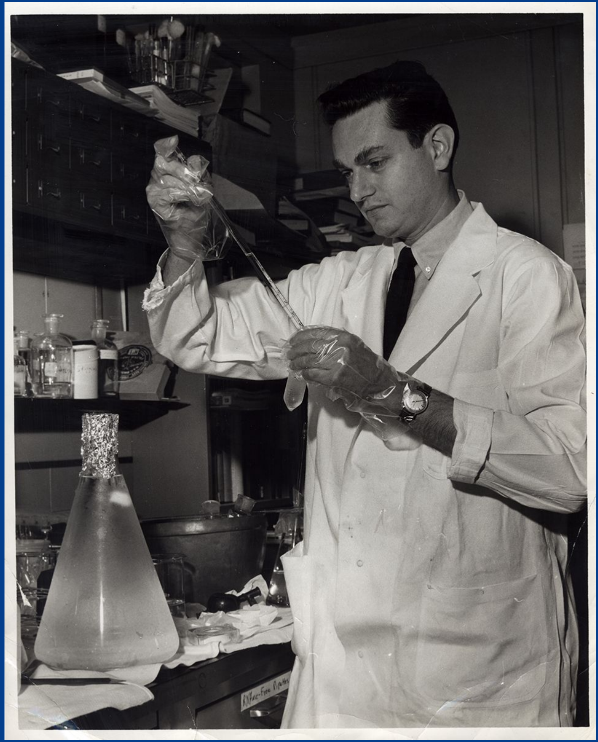
- Can pH be modeled as a function of time with simple features?
- Can fermentation stage be predicted reliably?
- Is the process predictable enough for real time user tools?

Evaluation: Experimental Setup

Libraries & Tools
- Google Colab
- Scikit-learn
- Pandas / NumPy
- Matplotlib
- ZoneInfo, datetime
- ipywidgets (interactive tool)

Hyperparameters
- 400 trees for both RF models
- max_depth = None
- Grouped jar split (culture vs no culture)

Evaluation Data
- Predictions vs actual pH
- Stage predictions vs stage labels
- Error analysis and visual graphs



Evaluation: Results

Regression Model (pH)

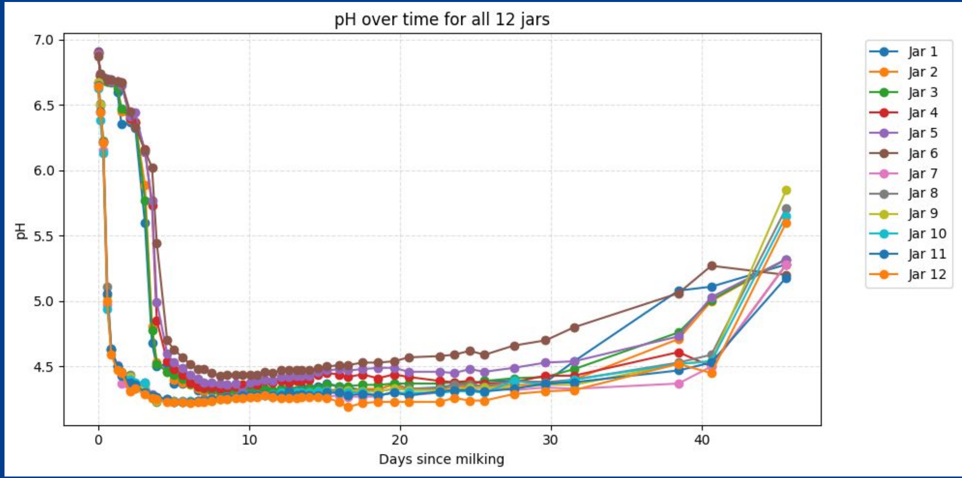
- RMSE ≈ 0.0375
- $R^2 \approx 0.9976$
- Closely tracks real pH curves

Classification Model (Stage)

- 100% test accuracy
- Perfect F1 score
- Correctly identified all stages

Observations

- Culture jars drop pH faster early
- Non-culture jars shift more slowly
- Interactive tool visualizes differences



Evaluation: Discussion

Key Findings

- Raw milk fermentation follows a predictable pattern
- Time + culture status explain most variation
- Stage labeling method worked well

Comparison to Literature

- Industrial studies don't model natural fermentation
- Spoilage models ignore intermediate stages
- pHase provides full-stage prediction

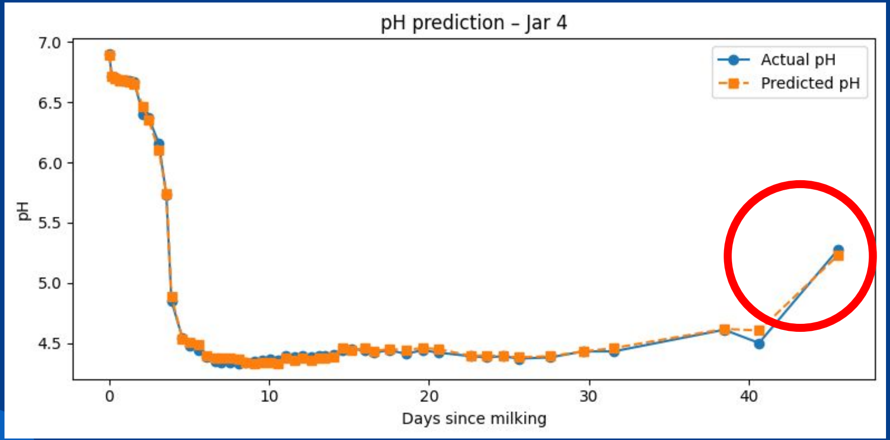
Additional Observations

- Jars lost ~20% volume during fermentation
- Culture adds significant speed early on


pHase - Demo

Limitations

- Limited to 12 jars, 1 milk source
- Only one temperature range (67–69°F)
- Late cheese stage data incomplete
- pH readings take 1–2 minutes per jar
- Time and resource constraints



Future Work & Conclusion

Conclusion

- ML can model raw milk fermentation accurately
- pH and stages follow stable patterns
- pHase provides real-time predictions

Future Work

- Use multiple farms and more jars
- Collect longer and more frequent measurements
- Improve stage labeling with direct observations
- Add bacterial testing
- Expand interactive tool into a public app

References

Aydogdu, M., Altuntas, A., & Kocak, P. (2023). pH: The fundamentals for milk and dairy processing – A review. Journal of Food Science and Nutrition, 11(7), 4567–4583. Retrieved from https://www.researchgate.net/publication/371901325

D'Amico, D. J., & Donnelly, C. W. (2010). Microbiological and chemical changes in raw milk during storage. Journal of Dairy Science, 93(3), 910–921. https://doi.org/10.3168/jds.2009-2761

Gagara, M., Nonga, H. E., & Ijumba, J. N. (2022). Hygienic quality of raw and fermented cow milk in the local market in Tanzania. Veterinary Medicine International, 2022, Article 9375221. https://doi.org/10.1155/2022/9375221

Tomovska, J., Gjorgievska, B., & Kuzmanova, S. (2016). Examination of pH, titratable acidity and antioxidant activity in fermented milk. Journal of Food and Nutrition Sciences, 4(3), 51–57. Retrieved from https://www.davidpublisher.com/Public/uploads/Contribute/58ca5d728d24d.pdf

Yang, H., Choi, J., Lee, S., & Kim, D. (2023). Investigation on bacterial growth and pH in milk after the introduction of a model spoilage bacterium. Microorganisms, 11(10), 2259. https://doi.org/10.3390/microorganisms11102259

SafeMilk Labs. (2022). Understanding the pH of milk and milk products: A key indicator of quality and freshness. Retrieved from https://safemilklabs.com/understanding-the-ph-of-milk-and-milk-products-a-key-indicatorof-quality-and-freshness/

University of Guelph. (2015). Raw milk quality. In Cheese-Making Technology eBook. Retrieved from https://books.lib.uoguelph.ca/cheesemakingtechnologyebook/chapter/raw-milk-quality/

Tetra Pak. (2023). Dairy processing handbook. Retrieved from https://dairyprocessinghandbook.tetrapak.com/chapter/chemistry-milk

SafeMilk Labs. (2022). Milk spoilage: Methods and practices of detecting milk quality. Retrieved from https://www.researchgate.net/publication/269963794_Milk_Spoilage_Methods_and_Prac tices_of_Detecting_Milk_Quality

Utah State University Extension. (2021). Tips to safely ferment at home. Retrieved from https://extension.usu.edu/preserve-the-harvest/research/tips-to-safely-ferment-at-home

Thanks!

Do you have any questions?
noberdier@oakland.edu

Please keep this slide for attribution