Napa Vine Advisory

An agentic decision-support tool that forecasts winegrape yield and quality using a climate-driven ML pipeline

Maria Romo Nichols
Advised by Gagan Garg

JOHNS HOPKINS
WHITING SCHOOL of ENGINEERING

Napa Vine Advisory
Climate-Informed Decisions

Introduction

Climate change is actively compressing and shifting growing seasons for wine grape growers, increasing vintage-to-vintage variability.

Small, independent Napa Valley vintners consistently make harvest timing and resource decisions oriented around these changing seasons, but **lack the same access to proprietary analytics platforms** that their larger competitors may have.

This project asks: can entirely **publicly available climate and agricultural data** produce reliable pre-season forecasts of grape ripeness (an indicator of grape quality) and yield (tons crushed)?

Objectives

1. **Forecast** harvest-season Brix and tons crushed for **Cabernet Sauvignon, Pinot Noir, and Chardonnay** using only free, public data sources

2. **Identify** which climate variables most influence **vintage quality and yield** through interpretable ML

3. **Translate** model outputs into **plain-language advisories** that are accessible to growers without data science backgrounds

Materials and Methods

Data sources (all public, free)
PRISM climate grids: 1981–present, 4 km res.
CIMIS weather stations: seasonal evapotranspiration (ETo) for Napa Valley
CDFA Crush Report: annual Brix and tons crushed by variety and district
NASS: bearing acreage by county and variety
SSURGO: soil texture and available water capacity

Feature engineering
Growing degree days (Apr–Oct), frost days, heat stress days (>95°F), winter precipitation, seasonal ETo, drought severity score, prior-year Brix and tons (lag-1)

Models
Baselines: historical mean, persistence (prior year), Winkler GDD linear regression, full OLS
Primary: Elastic net with Δ-targets (predict year-over-year change, not absolute level)
Comparison: LightGBM multi-output gradient boosting with SHAP TreeExplainer

Evaluation: Walk-forward cross-validation, 2005–2024

Results or Findings

1 Discoveries

- **Model works for brix (ripeness):**
  - **Beats every baseline including the Winkler GDD model (industry standard since 1944)**
  - Cabernet Sauvignon error cut nearly in half vs. historical mean (RMSE 0.59 vs. 1.07° Brix)
  - Chardonnay predicted within ±0.33° Brix — inside harvest decision precision (±1° threshold)
  - **SHAP confirms real viticultural signals:** growing degree days, heat stress days, and veraison Tmax are top drivers across all three varieties

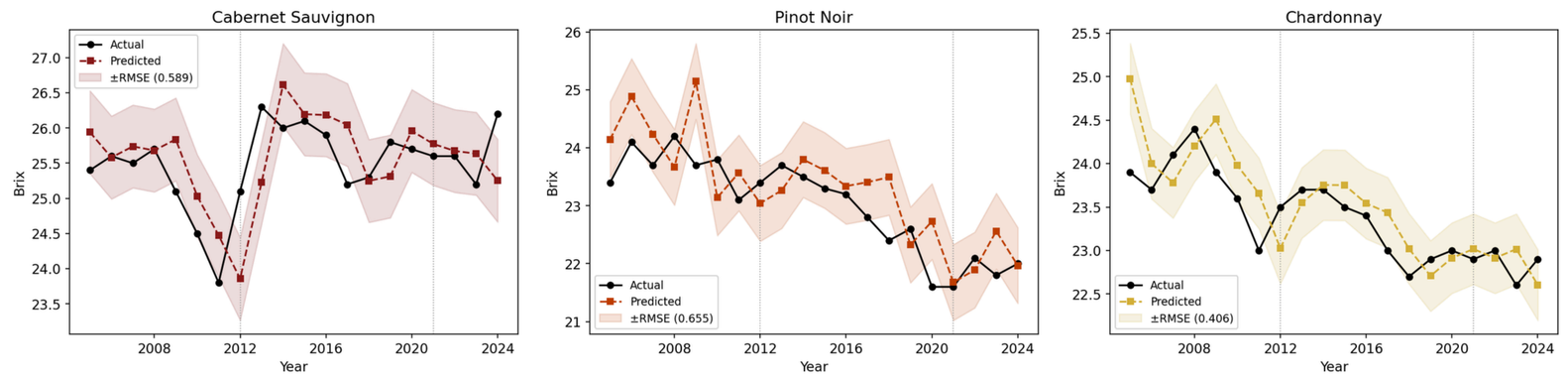
- **Model falls short for tons crushed (yield):**
  - Negative R^2 across all models and all varieties
  - No variation of the model beats a naïve guess
  - Crush volume is driven, and often pre-determined, by replanting cycles and contract pricing, not weather

Model comparison table (Brix RMSE where lower is better):

Variety	Historical Mean	Winkler GDD	Elastic Net Model
Cabernet Sauvignon	1.07	0.96	0.59
Pinot Noir	0.94	0.94	0.65
Chardonnay	0.51	0.53	0.41

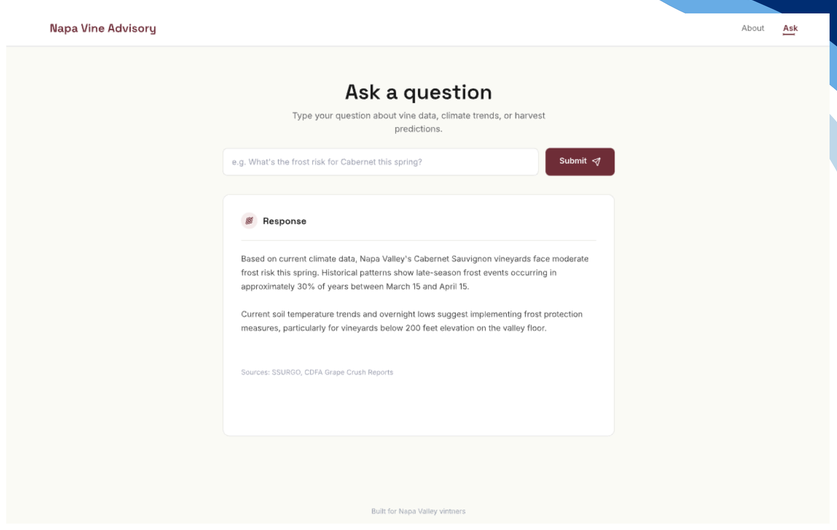
2 Performance

This free, reproducible climate pipeline, built exclusively on public data, forecasts Napa Valley grape ripeness within harvest-decision precision, **outperforming the 80-year-old industry benchmark**.



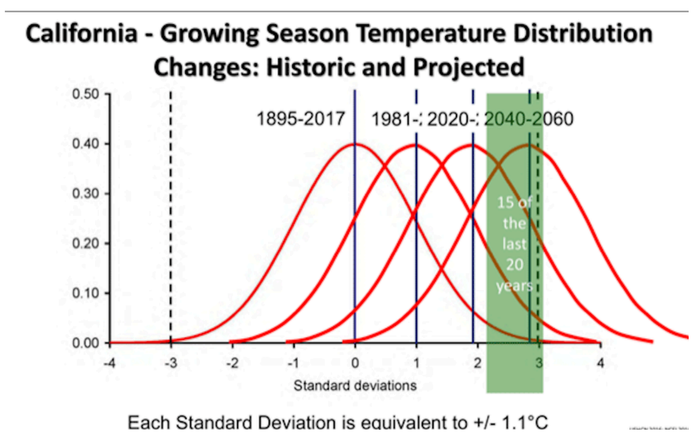
3 Translation

- Model outputs (predicted Brix delta, top SHAP drivers) are passed to an **LLM agentic advisory layer**, including a base agent, RAG agent, and an evaluator agent.
- **Translates numeric forecasts into a 2–3 sentence plain-language recommendation** tailored to the variety and vintage conditions
- Simple interface designed for growers without data science backgrounds, enhancing this project's emphasis on accessibility



4 Why?

- **Napa Valley has warmed ~2.5°F since 1950**
- Extreme heat events (>95°F) during veraison have increased in frequency, directly **degrading sugar-acid balance and reducing grower control** over harvest timing
- The top 10 Napa wineries control the majority of valley acreage
- Small, independent vintners represent the **cultural backbone of the region** but operate **without analytics infrastructure**



Brooks, L. "Climatologists Say Cabernet's Days as King in Napa are Numbered: How the Changing Climate Could Create Unfavorable Conditions for This Particular Grape."

Conclusion

Climate-driven Brix forecasting is achievable with free, public data. This model outperforms an 80-year-old industry benchmark and delivers results via a plain-text, interactive interface.

As rising temperatures threaten Napa's signature varieties, **publicly-accessible analytics tools** represent a low-cost path of resilience for small, independent vintners who define the valley's culture.