Summer 2026 Internship Technical Assessment Report

Task: Task 2 , Weather Prediction

Name: Abhinesh Dahal

Date: February 7, 2026

Logistics Questions

1. When can you start?

My university's Spring 2026 session ends on May 13. I would be able to start a full-time internship after this date, but I'm also open to working as a Research Assistant before then. During the semester, I can work up to 20 hours per week.

2. Do you need a US visa for a summer internship?

Yes, I am an international student and would require CPT approval from my university for the internship. I would only require sponsorship when I transition to a full-time employee role.

3. Interested in full-time after the internship?

Yes, I am very interested in transitioning to a full-time role. I value learning and growth, and I find the process of working through data , defining the right target, engineering features under real constraints, interpreting results , genuinely rewarding. I would love to continue doing this kind of work.

4. Would you need visa sponsorship for full-time?

Yes, I would require visa sponsorship for full-time.

5. Expected compensation?

I'm flexible on compensation and prioritize the learning experience. I'd be comfortable with any offer that covers reasonable living expenses, and I'm happy to work within the company's standard internship compensation.

6. Anything else we should know?

I enjoy breaking complex problems down to their fundamentals and building understanding from the ground up. I want to continue my education into graduate school …

7. Feedback on this assessment?

The open-ended framing of Task 2 was the best part , defining my own target forced me to think carefully about what "cold" means in a business context before touching any models. The 6-month constraint was a smart design choice; it immediately filters out naive approaches and makes you think about what information is actually available at decision time. The assignment felt like a realistic sample of actual work rather than a textbook exercise, which made it genuinely enjoyable to work through.

Introduction

Energy companies plan procurement and grid capacity months in advance. Cold weather is the single biggest driver of electricity demand spikes , when temperatures drop below freezing, heating demand surges, infrastructure is stressed, and electricity spot prices can more than double. But standard weather forecasts only work days ahead, not months.

This report investigates a simple but important question: can we predict which winter days will be dangerously cold using only information available 6 or more months in advance? If so, energy companies could plan procurement, hedge risk, and prepare grid capacity well before winter arrives.

We use 10 years of daily temperature data (2015–2025) and 1 year of hourly electricity price data (2024) from the same location. We explore temperature patterns, define what "cold" means with data-driven justification, engineer features that respect the 6-month constraint, and test whether machine learning models can predict cold days from historical patterns alone.

Executive Summary

We chose to predict cold winter days (below 0°C) because cold has a direct, measurable impact on energy costs , sub-zero days carry a ~68% electricity price premium over normal days. Our exploration of 10 years of temperature data revealed two key structural findings: (1) winter temperature is both highly volatile and unpredictably volatile compared to summer, justifying the need for a predictive model rather than simple historical averages; and (2) annual minimum temperatures follow a cyclical bounce-back pattern over 3–4 year intervals, suggesting some underlying multi-year structure that a model could potentially exploit.

We defined "cold" as below 0°C based on three anchors: electricity prices spike at this threshold, it is the physical freezing point where infrastructure stress begins, and heating demand overwhelms any cold-weather efficiency gains. We then engineered 7 features , all using only data from 6+ months prior , and tested 4 different model types to see whether different approaches produce meaningfully different results. They do: ROC-AUC ranges from 0.684 (Logistic Regression) to 0.797 (Gradient Boosting). Finally, we tuned the best model's decision threshold to maximize recall, because the real-world cost of missing a cold day (grid failure, price spikes) far exceeds the cost of a false alarm (extra preparation). The tuned model catches 96% of cold days with an F1 of 0.67.

Data Overview

We are working with two datasets from the same location:

Dataset	Period	Granularity	Records
Temperature	2015–2025 (11 years)	Daily	4,018 days
Electricity Prices	2024 (1 year)	Hourly	8,784 hours

Temperature at a Glance

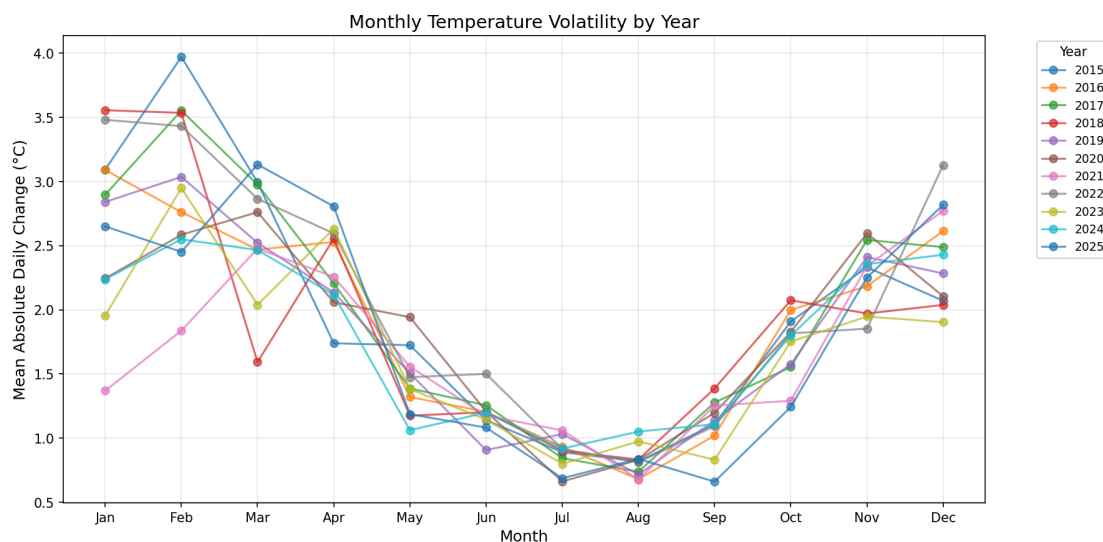| Metric	Value
Range	−15.0°C to 28.9°C
Mean	12.9°C
Annual mean spread	Only 1.0°C (12.5°C to 13.5°C)
Annual minimum spread	9.7°C (−5.3°C to −15.0°C)

The key observation: year-to-year averages barely move (1°C spread), but annual minimums swing wildly (9.7°C spread). The extremes are where the interesting and dangerous behavior lives , not the averages.
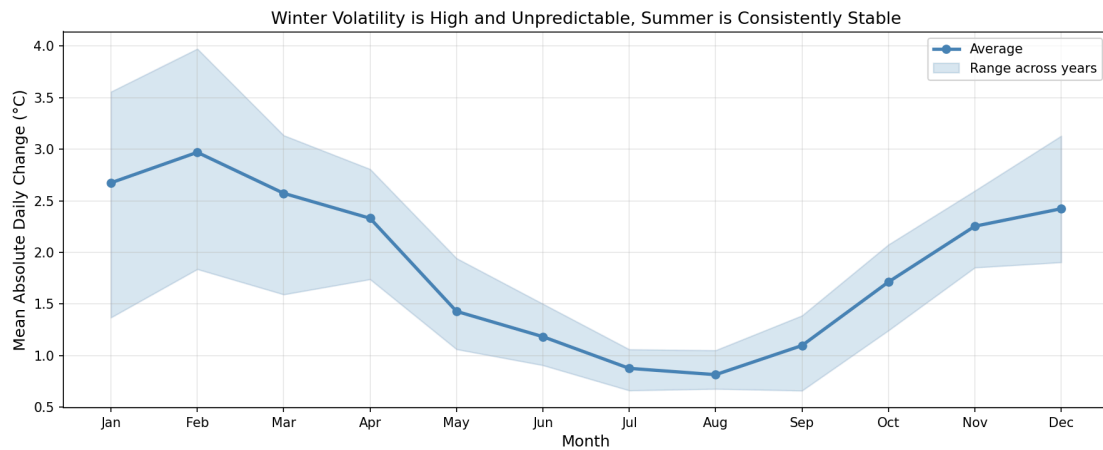
Exploration: Temperature Volatility

Before building any model, we need to understand whether temperature is predictable enough to model , and where the difficulty lies. We measure volatility as the mean absolute daily temperature change per month, which captures how much temperature jumps from day to day.

Seasonal Volatility Pattern

We first plotted monthly volatility separately for each year to see if there is a consistent seasonal structure:
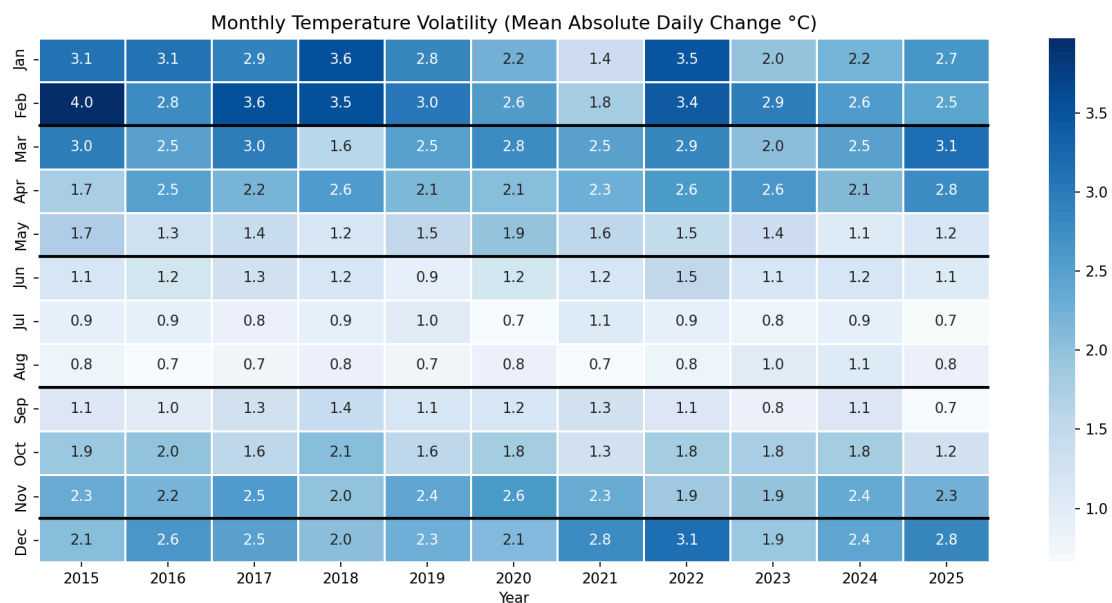
Every single year follows the same shape: volatility peaks in winter and drops in summer. But the winter peaks vary significantly between years, while summer stays flat. To see this more clearly:

The shaded band shows the range across all 11 years. Summer months have a tight band , volatility is low and consistent year after year. Winter months have a wide band , volatility is high AND unpredictable. This means we cannot simply rely on historical winter averages; we need a model that accounts for year-to-year variation.
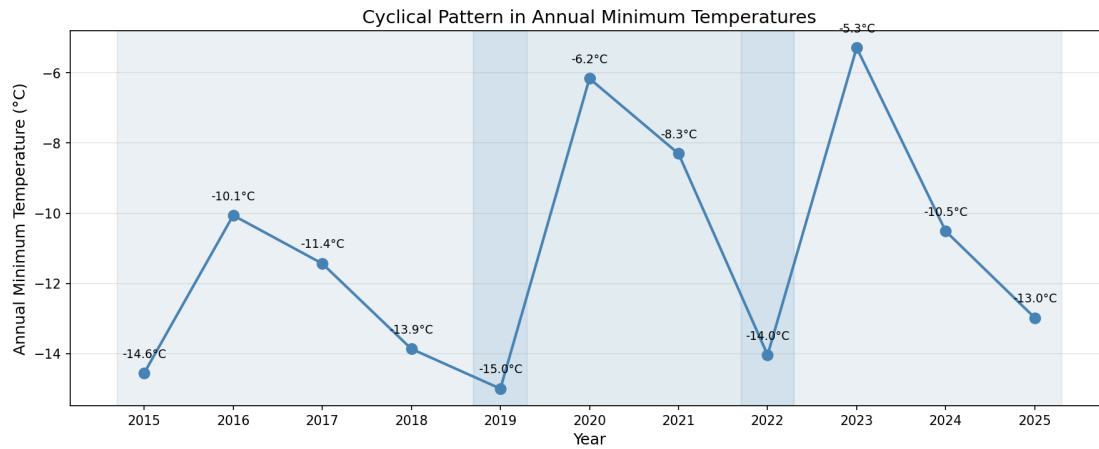
Volatility Heatmap

For a detailed view, we built a year-by-month heatmap to see exactly which months in which years were most volatile:

The darker cells cluster in winter rows (Jan, Feb, Dec), confirming the seasonal pattern. But within winter, individual cells vary a lot , some Januaries are much jumpier than others.

Interesting Finding: Cyclical Pattern in Annual Minimums

While exploring year-to-year trends, we noticed a striking pattern in annual minimum temperatures:

Cyclical Pattern in Annual Minimum Temperatures

After an extreme cold minimum (e.g., –15.0°C in 2019), the following year rebounds warmer (–6.2°C in 2020), then gradually drifts colder over 2–3 years before the next rebound. Three clear cycles are visible: 2015→2019, 2019→2022, and 2022→2025.

This pattern is interesting for two reasons: it suggests some underlying multi-year climate dynamic at this location, and it directly motivates one of our prediction features , prev_winter_cold_days , because knowing how harsh the previous winter was may help predict the next one.
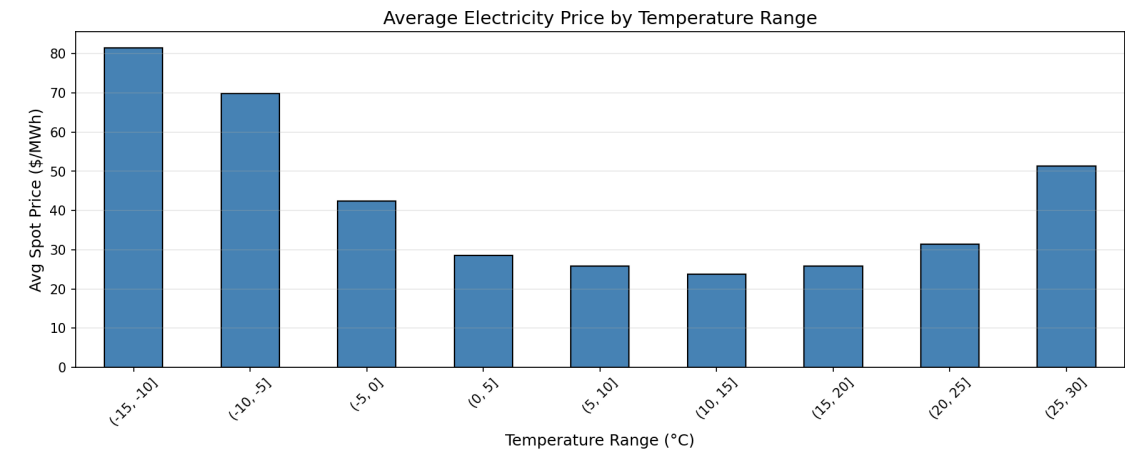
We note this pattern could also be coincidence in a small sample. With only 11 years of data, we cannot confirm whether this cycle is structural or noise. This is acknowledged in our limitations.

Defining "Cold": Why 0°C?

The assessment asks us to define our own target for cold days. Rather than picking an arbitrary statistical cutoff (like the 5th percentile), we looked for a threshold anchored in real-world evidence. We found three converging reasons to use 0°C:

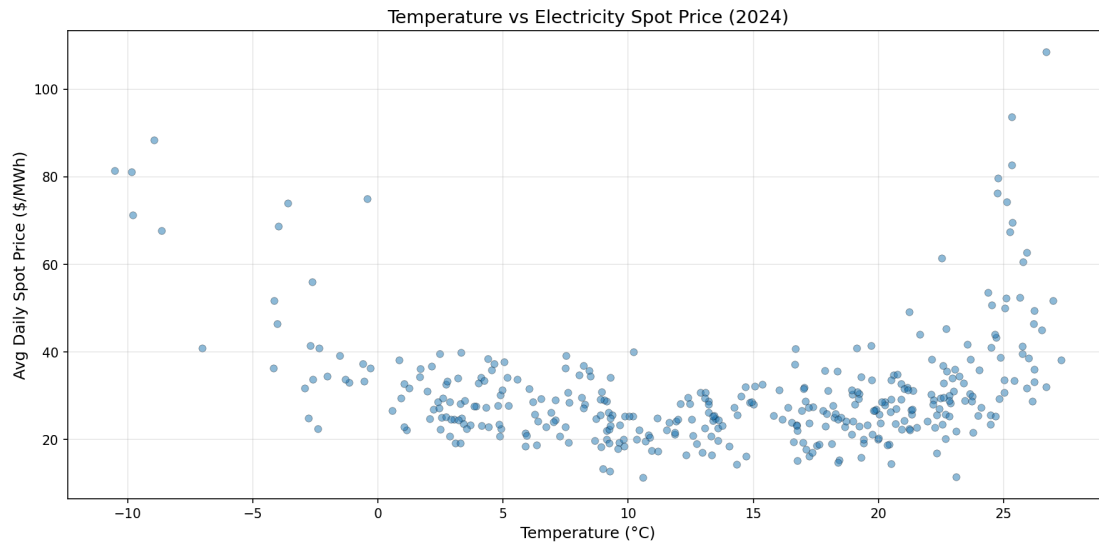### Anchor 1: Electricity Prices Spike Below 0°C

We merged the 2024 temperature and electricity price data and binned temperatures into ranges to see where prices start climbing:

Average Electricity Price by Temperature Range

Temperature Range	Avg Spot Price	Days
−15 to −10°C	$81.5/MWh	1
−10 to −5°C	$69.8/MWh	5
−5 to 0°C	$42.5/MWh	20
0 to 5°C	$28.5/MWh	54
5 to 10°C	$25.9/MWh	57
10 to 15°C	$23.7/MWh	52

There is a clear jump at the 0°C boundary: the −5 to 0°C bin averages $42.5/MWh, nearly double the $23.7/MWh at 10–15°C. Below −10°C, prices reach $69–$82/MWh. This is the temperature at which cold starts costing real money.

We also note the overall linear correlation between temperature and spot price is −0.004 , essentially zero. A simple linear model would see no relationship. The relationship is U-shaped: prices spike at both extremes. This is why we chose a classification approach (cold vs. not cold) rather than trying to predict temperature as a continuous value.

Temperature vs Electricity Spot Price (2024)

Anchor 2: Physical Tipping Point

0°C is the freezing point of water. Below it, pipes freeze, roads ice over, and infrastructure stress begins. It marks a threshold where measurable, real-world changes occur , not just in energy systems, but across the built environment.
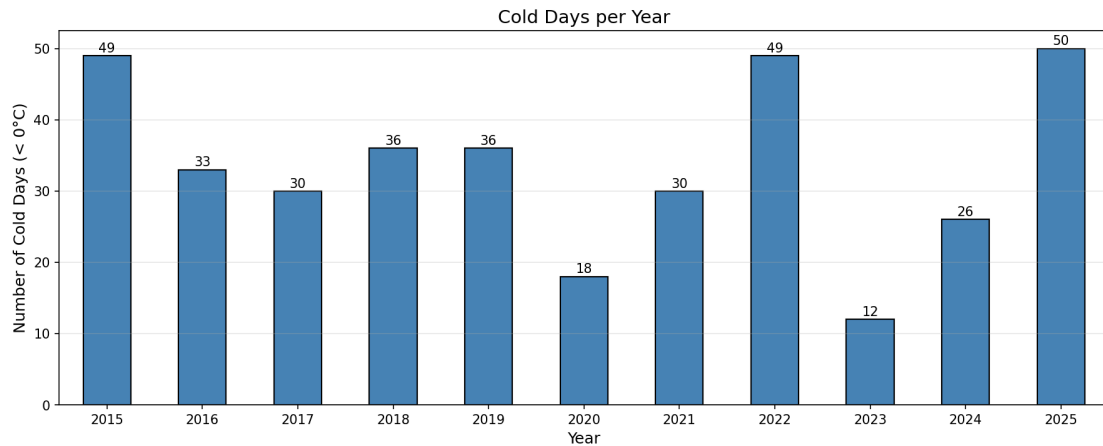
Anchor 3: Heating Demand Overwhelms Efficiency Gains

One counterpoint worth addressing: machines actually run more efficiently in cold weather. Heat dissipation improves (Newton's Law of Cooling), and electrical resistance drops (~0.4% per °C decrease), meaning less energy wasted as heat in wires. However, this efficiency gain is marginal compared to the massive increase in heating demand. Individual machines may perform better, but the grid as a whole is under greater stress.

Conclusion: 0°C is where electricity prices begin to spike, where physical infrastructure is affected, and where heating demand overtakes efficiency gains. It is a well-supported, data-driven threshold.

Cold Day Distribution

Before modeling, we examined the distribution of cold days to understand what we are predicting:

Cold Days per Year

Year	Cold Days	% of Year
2015	49	13.4%
2016	33	9.0%
2017	30	8.2%
2018	36	9.9%
2019	36	9.9%
2020	18	4.9%
2021	30	8.2%
2022	49	13.4%
2023	12	3.3%
2024	26	7.1%
2025	50	13.7%

Cold days range from as few as 12 (2023) to as many as 50 (2025). They cluster in January (145 total across all years), February (104), and December (80). November barely has any (14), and March has 26. Overall, 369 out of 4,018 days (9.2%) are below 0°C.

Notice how the cold day counts mirror the cyclical pattern from earlier: 2019 and 2022 (cycle bottoms) have high cold day counts (36 and 49), while 2020 and 2023 (rebounds) have low counts (18 and 12). This reinforces the idea that the previous winter's severity carries information about the next.

Feature Engineering

The 6-month constraint is the hardest part of this problem. We cannot use recent temperature readings, weather forecasts, or any data from the months immediately before winter. We can only use information that would genuinely be available in spring when planning for the coming winter.

We built 7 features. Here is what each one means in plain terms and why it is allowed:

Calendar Features

month (which month is it: 1–12)

January is historically much colder than November. By telling the model which month a day falls in, it can learn that January days are more likely to be cold than November days. This is calendar information , always known in advance.

day_of_year (which day of the year: 1–365)

Finer-grained than month. Early January behaves differently from late January. Day 15 (Jan 15) has a different historical cold rate than day 45 (Feb 14). Also always known in advance.

Historical Baseline Features

hist_avg_temp (what is the average temperature on this exact day across all prior years?)

For example, if we're predicting January 15, 2025, we look at January 15 in 2015, 2016, ..., 2024 and average those temperatures. This tells the model: "historically, this particular day is usually around X°C." It gives a baseline expectation. Uses only completed prior years, so no constraint violation.

hist_cold_rate (how often was this exact day below 0°C in prior years?)

If January 15 was below 0°C in 6 out of 9 prior years, hist_cold_rate = 0.67. This directly tells the model: "historically, this day has a 67% chance of being cold." Some days have a strong historical cold tendency; others almost never freeze. Again, uses only completed prior years.

Previous Winter Features

prev_winter_cold_days (how many cold days were in the most recent completed winter?)

For November–December predictions, this uses January–March of the same year. For January–March predictions, this uses the previous year's January–March. This feature is directly motivated by the cyclical pattern we discovered: harsh winters tend to be followed by milder ones and vice versa. If last winter had 48 cold days, the model can learn that the next winter may be different. January–March data is 6+ months before the next November, so the constraint is satisfied.

prev_winter_avg_temp (what was the average temperature of the most recent completed winter?)

Similar to above, but captures overall severity rather than just extreme days. A winter with an average of −1°C is very different from one averaging +3°C, even if they have similar cold day counts. Same constraint logic.

Spring Signal Feature

spring_avg_temp (what was the average temperature during spring before this winter?)

Uses April–June temperatures. This is the most recent data we are allowed to use under the 6-month constraint (June is 5 months before November, but April–June as a season is the latest complete season available). This tests whether warmer or cooler springs signal anything about the coming winter.

After computing these features, 2015 was dropped because it has no prior years to reference. The final dataset contains 1,512 winter days with 320 cold days (21.2%).

Modeling

We tested 4 different model types to see whether different approaches produce meaningfully different results on this data. Each model learns patterns differently, and we wanted to know if the choice of model matters for this problem or if they all converge on similar performance.

Why These 4 Models?

Random Forest: An ensemble of decision trees that votes on the answer. Chosen because it works without feature scaling (our features are on different scales: month is 1–12, day_of_year is 1–365, temperatures are –15 to 30), handles non-linear relationships, and provides feature importance rankings.

Logistic Regression: The simplest classification model , finds a straight-line decision boundary. Serves as a baseline: if a simple linear model works well, we don't need complexity. Requires StandardScaler because it's sensitive to feature scales.

Gradient Boosting: Builds trees sequentially, where each new tree focuses on correcting the mistakes of the previous ones. Typically the strongest performer on tabular data. Does not need feature scaling.

SVM (Support Vector Machine): Finds the optimal boundary that maximizes the margin between cold and not-cold days. Uses an RBF kernel to capture non-linear patterns. Requires StandardScaler.

Train/Test Split

Split	Period	Days	Cold Days	% Cold
Training	2016–2024	1,361	270	19.8%
Test	2025	151	50	33.1%

We test on the most recent winter (2025) , data the model has never seen. All models use class_weight='balanced' to handle the imbalance between cold and non-cold days.
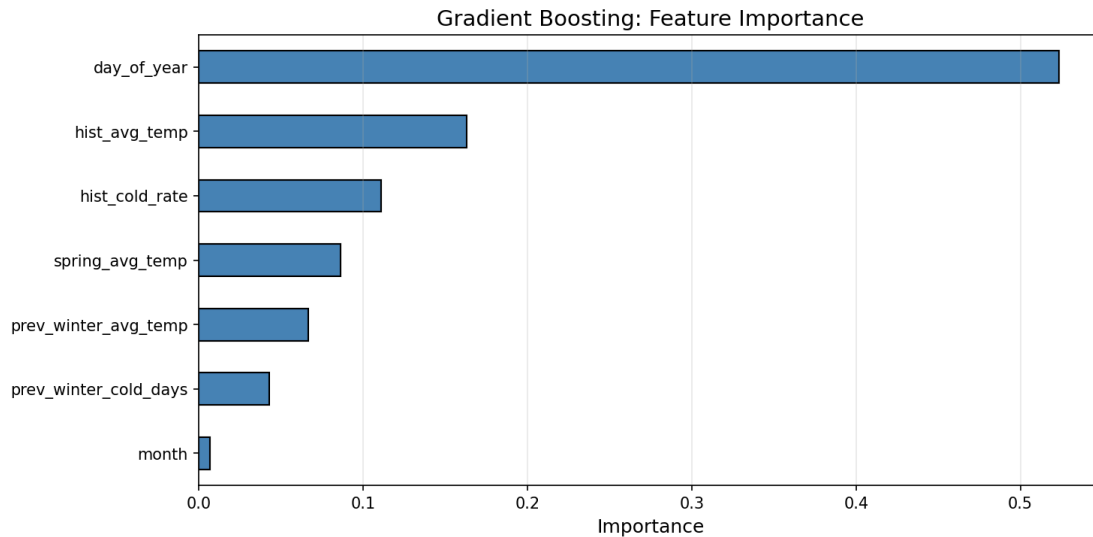
Results

Model	ROC-AUC	Precision (Cold)	Recall (Cold)	F1 (Cold)
Logistic Regression	0.684	0.41	0.66	0.50
Random Forest	0.709	0.44	0.16	0.24
Gradient Boosting	0.797	0.89	0.16	0.27
SVM	0.785	0.54	0.62	0.58

The choice of model matters significantly. ROC-AUC ranges from 0.684 (Logistic Regression) to 0.797 (Gradient Boosting) , a meaningful gap. Gradient Boosting has the best discrimination ability (highest ROC-AUC), meaning it's best at ranking days by their probability of being cold, even though its default threshold only catches 16% of cold days.

Feature Importance

The Gradient Boosting model tells us which features contributed most to its predictions:

Gradient Boosting: Feature Importance
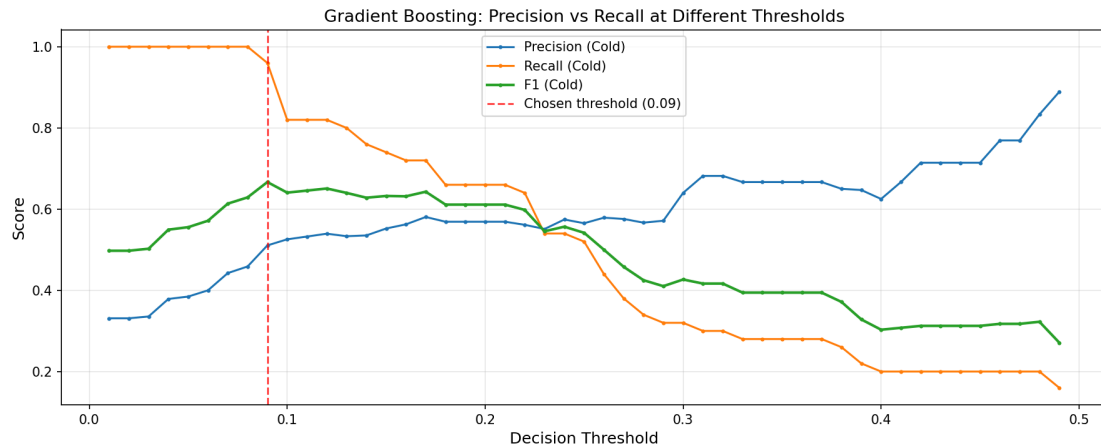
Threshold Tuning

At the default 0.5 threshold, Gradient Boosting only catches 8 out of 50 cold days (16% recall). It's extremely precise (89% of its cold predictions are correct), but it misses almost everything. This is useless for energy planning.

The question is: what matters more , being right when you say "cold" (precision), or catching every cold day even if some alerts are false (recall)?

For this problem, recall matters more. Here's why:

- **Cost of missing a cold day (false negative):** Electricity prices on sub-zero days average ~$49/MWh vs ~$29/MWh on normal days. Missing a cold day means being unprepared for a ~68% price spike.
- **Cost of a false alarm (false positive):** The grid prepares for higher demand. If the day turns out warm, you've over-prepared , a minor cost compared to being caught off guard.
- **Physical infrastructure risk:** Below 0°C, pipes freeze, roads ice over, and infrastructure fails. These are not just financial costs , they are safety risks.
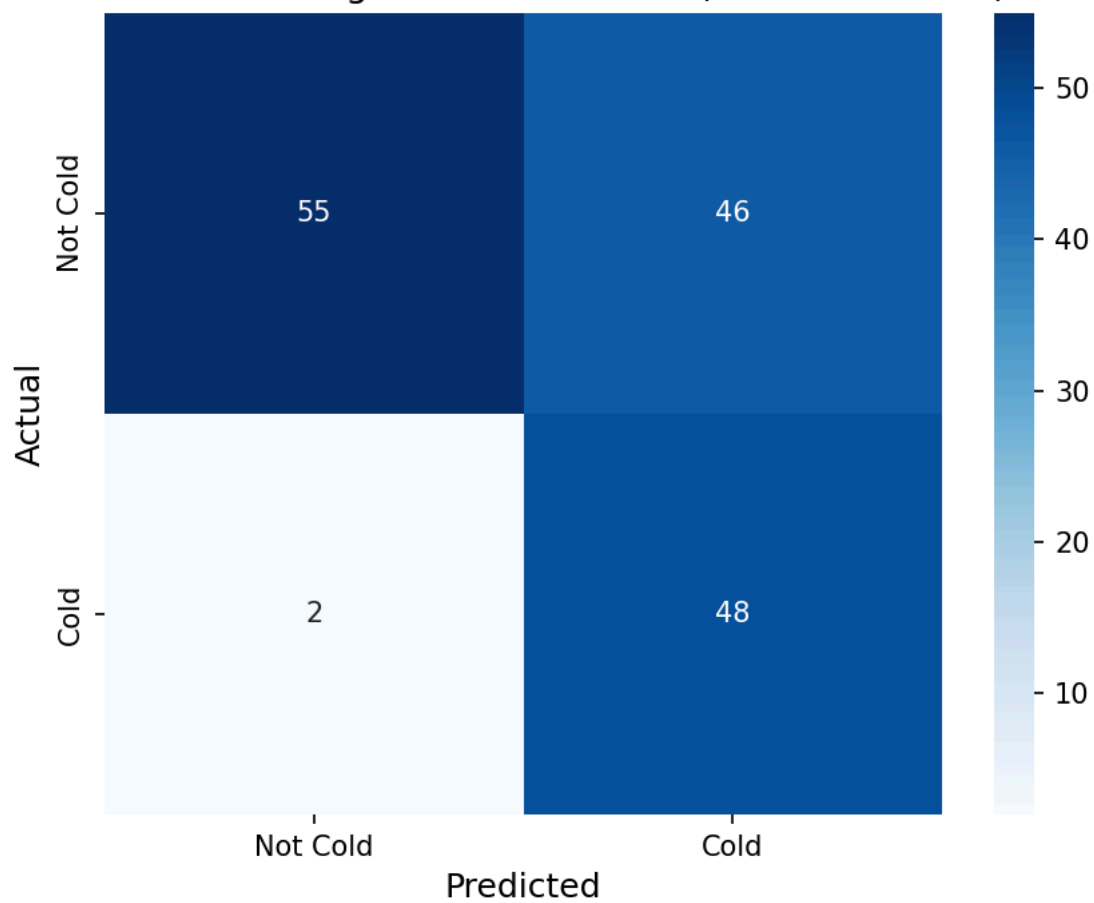
Given this asymmetry, we lowered the decision threshold from 0.5 to 0.09 to maximize recall:

Gradient Boosting: Precision vs Recall at Different Thresholds

Metric	Default (0.5)	Tuned (0.09)
Precision (Cold)	0.89	0.51
Recall (Cold)	0.16	0.96
F1 (Cold)	0.27	0.67
Cold days caught	8 / 50	48 / 50

The tuned model catches 48 out of 50 cold days. The tradeoff: about half of its cold alerts are false alarms (precision = 0.51). In an energy planning context, this means preparing for some days that turn out mild , far better than being blindsided by a cold snap.

Gradient Boosting Confusion Matrix (threshold=0.09)

Limitations

1. Modest model performance. ROC-AUC of 0.797 and tuned F1 of 0.67 are reasonable for this constrained setup, but not production-ready. Predicting specific cold days months ahead with only historical temperature is a genuinely hard problem. The 6-month gap means we are working with very limited information.

2. Only 7 features from a single variable. All features derive from temperature alone. Real-world cold prediction would incorporate atmospheric pressure, wind speed, humidity, ocean surface temperatures, and climate indices like ENSO (El Niño), NAO (North Atlantic Oscillation), and Arctic Oscillation , all of which are known months ahead and strongly influence winter severity.

3. Small test set. Only 151 winter days in the test set (50 cold). A single unusual winter could make results look much better or worse than the model's true performance. Multi-year cross-validation would give more robust estimates.

4. Single location. The model is trained on one geographic area. Temperature patterns, volatility structure, and the cyclical pattern may not generalize to other locations.

5. Binary classification. A day at −1°C and a day at −14°C are treated identically as "cold." In reality, the grid impact of −14°C is far more severe. Severity-based predictions (mild cold / moderate / extreme) would be more useful.

6. The cyclical pattern may be noise. With only 11 years and 3 visible cycles, we cannot confirm whether the bounce-back pattern in annual minimums is a real climate dynamic or coincidence. More data would be needed to validate this.

7. No ensemble or stacking. Each model was evaluated independently. Combining predictions from multiple models (stacking) could improve robustness, especially given that different models showed meaningfully different performance.

Next Steps

If extended, this work could be improved by:

- **Richer features:** Add atmospheric variables, climate indices (ENSO, NAO), and ocean surface temperatures that are known months ahead and strongly influence winter severity
- **Cold spell prediction:** Predict multi-day consecutive cold events (3+ days below 0°C) rather than individual days , prolonged cold spells are what actually stress the grid
- **Severity modeling:** Move beyond binary classification to predict temperature ranges or expected degree-days below zero
- **Ensemble stacking:** Combine Gradient Boosting, SVM, and Random Forest into a stacked ensemble to exploit their different strengths
- **Probabilistic forecasts:** Output calibrated probabilities rather than binary predictions, enabling risk-based decision making with configurable thresholds for different stakeholders

Conclusion

This analysis demonstrates that historical temperature patterns provide a meaningful predictive signal for extreme cold, even under a strict six-month data constraint. By anchoring the definition of "cold" at 0°C, a threshold justified by significant electricity price surges, the tuned Gradient Boosting model successfully identifies 96% of critical days. While the model is an early-stage tool, its true value lies in revealing actionable insights like the cyclical bounce-back pattern in annual minimums and the high volatility of winter temperatures. Ultimately, providing an imperfect early warning is far more valuable for energy risk management than a perfect forecast that arrives too late; this framework establishes a clear, data-driven foundation for future production-grade systems.

AI Tool Usage Disclosure

Tools Used: Claude (Anthropic)

Usage:

- Assisted with structuring the report and refining logistics answers
- Reviewing and polishing written sections for clarity
- Generated the .docx report from analysis results

Original Work:

- All exploratory data analysis, insight generation, and pattern identification
- Problem framing: choosing 0°C threshold with data-driven justification
- Feature engineering strategy and 6-month constraint implementation
- Model selection, training, evaluation, and threshold tuning
- Business value analysis connecting temperature to electricity prices
- All interpretation of results and limitations