ARROCHA & ASSOCIATES

Climate-Enhanced Pricing

Using Gradient Boosting Machines to Personalise Life Insurance Rates

IDS Conference — London

June 2025

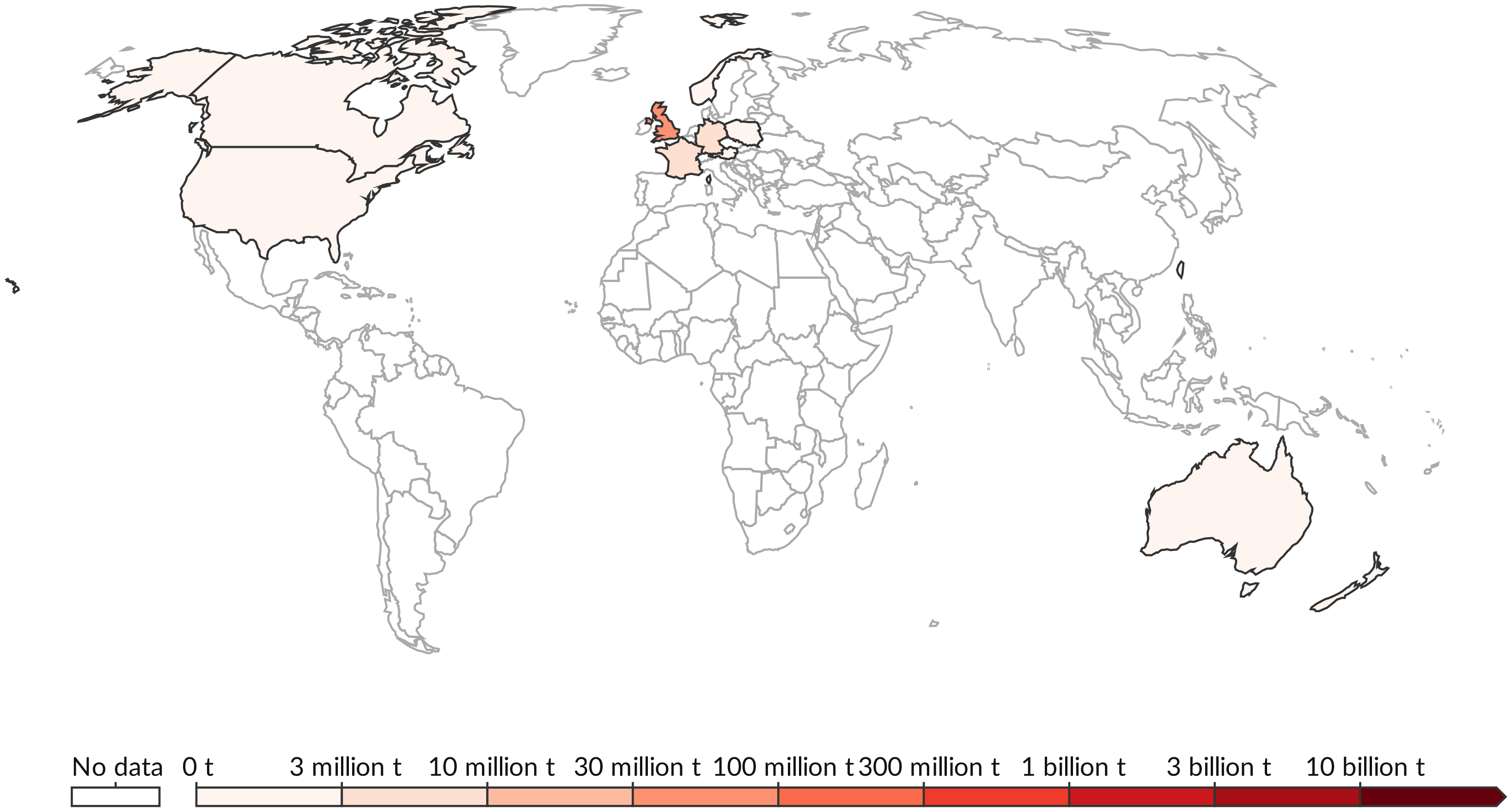
Carlos Arocha

Agenda

Problem Statement & Motivation

Annual CO_2 emissions, 1823

Carbon dioxide (CO_2) emissions from fossil fuels and industry[1]. Land-use change is not included.



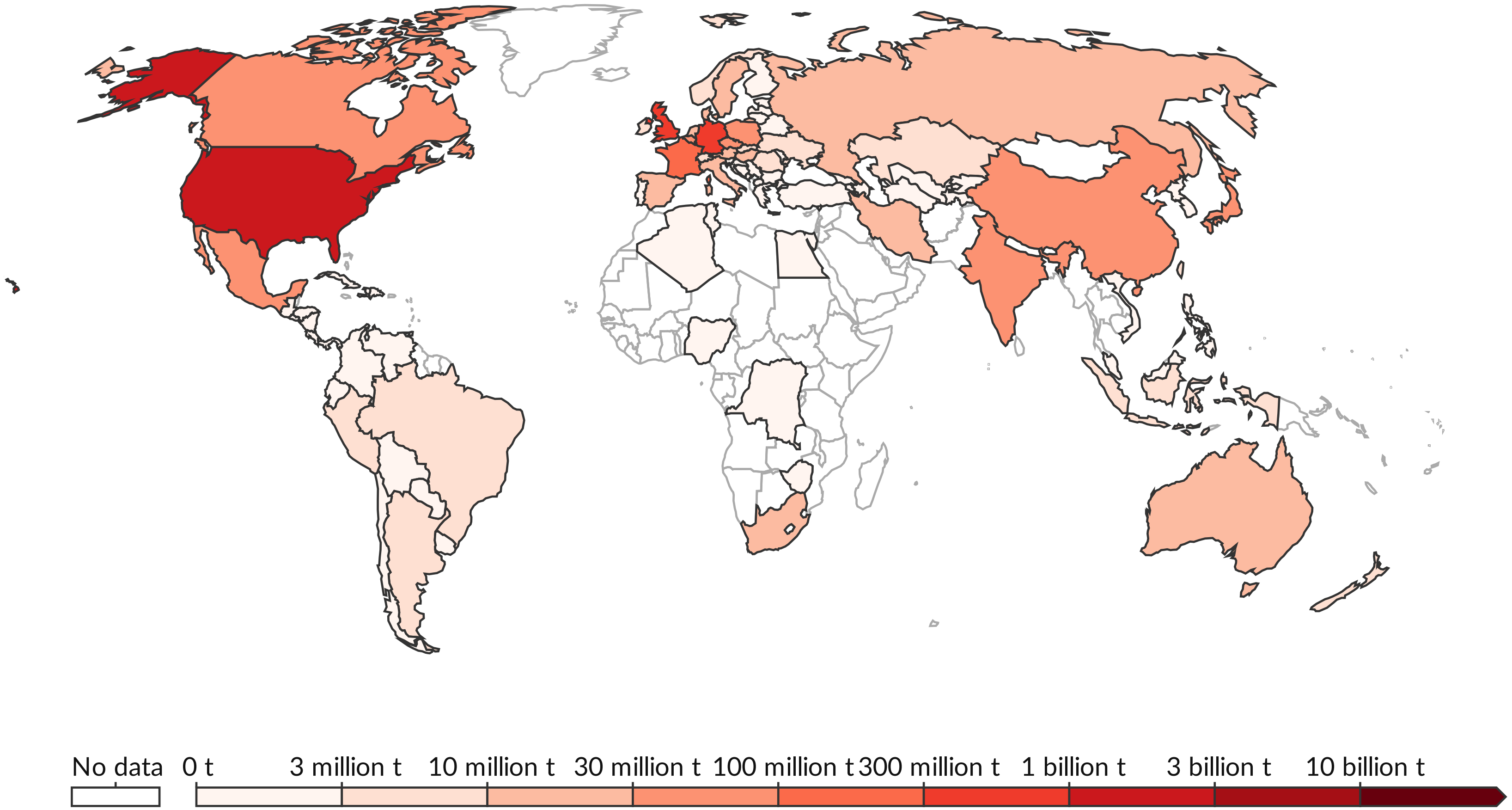
Data source: Global Carbon Budget (2024)

1. Fossil emissions Fossil emissions measure the quantity of carbon dioxide (CO_2) emitted from the burning of fossil fuels, and directly from industrial processes such as cement and steel production.
Fossil CO_2 includes emissions from coal, oil, gas, flaring, cement, steel, and other industrial processes.
Fossil emissions do not include land use change, deforestation, soils, or vegetation.

Annual CO_2 emissions, 1923

Carbon dioxide (CO_2) emissions from fossil fuels and industry[1]. Land-use change is not included.

Our World in Data



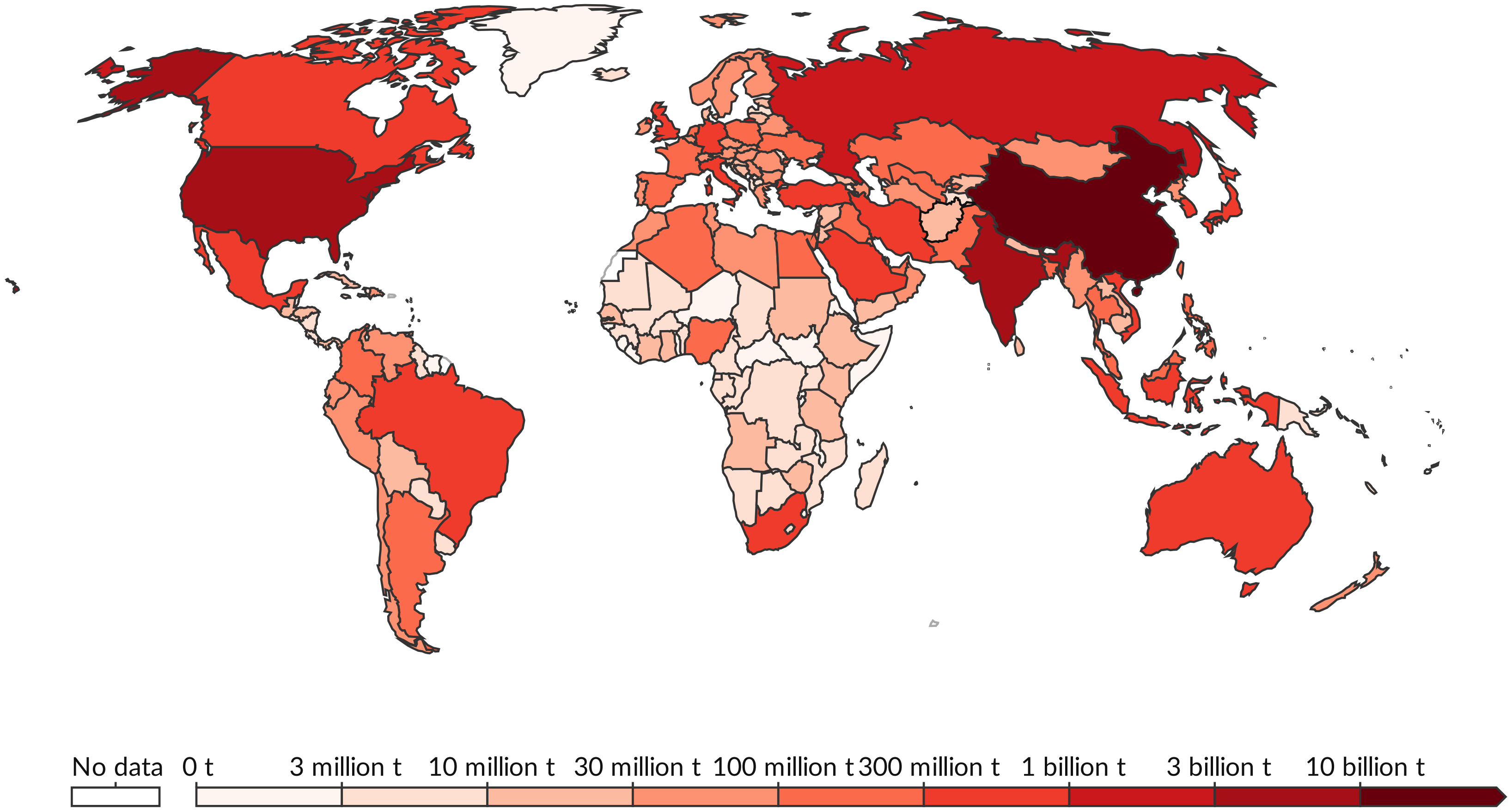
Data source: Global Carbon Budget (2024)

OurWorldinData.org/co2-and-greenhouse-gas-emissions |

1. **Fossil emissions** Fossil emissions measure the quantity of carbon dioxide (CO_2) emitted from the burning of fossil fuels, and directly from industrial processes such as cement and steel production.
Fossil CO_2 includes emissions from coal, oil, gas, flaring, cement, steel, and other industrial processes.
Fossil emissions do not include land use change, deforestation, soils, or vegetation.

Annual CO_2 emissions, 2023

Carbon dioxide (CO_2) emissions from fossil fuels and industry[1]. Land-use change is not included.



Data source: Global Carbon Budget (2024)

1. Fossil emissions Fossil emissions measure the quantity of carbon dioxide (CO_2) emitted from the burning of fossil fuels, and directly from industrial processes such as cement and steel production.
Fossil CO_2 includes emissions from coal, oil, gas, flaring, cement, steel, and other industrial processes.
Fossil emissions do not include land use change, deforestation, soils, or vegetation.

ERM risk classes impacted by climate-related risks

Risk Class	Physical Risks	Transition Risks	Legal / Reputation Risks
Market	Medium	High	High
General Insurance	High	Medium	High
Longevity	Medium	Low	Low
Mortality/Morbidity	Medium	Low	Low
Lapse	Low	Medium	Low
Counterparty	Medium	High	Medium
Operational	Low	Medium	Low
Strategic	Medium	High	High
Reputational	Low	Medium	High

Source: International Actuarial Association, Climate Risk Task Force (2020), *Importance of Climate-Related Risks for Actuaries*

How actuarial work is exposed to climate-related change

Climatic Impacts		Socio-Economic Impacts		Impacts on Actuarial Work
Direct	**Indirect**	**Social**	**Economic**	• Changes to modelling and assumptions • Development of products including re-design, pricing, exclusions, etc. • Changes to risk management practices • Changes to capital management practices • Prevised/new investment management practices • Changes to financial stability management • Disclosure that allows for climate risk • Broader application of actuarial work
• Heatwaves • Storms • Floods • Sea level rise • Bush fires • Droughts	• Air pollution • Water and food supply • Diseases	• Migration • Health infrastructure • Emergency and social services • Consumer behavior	• GDP growth • Investor preferences • Infrastructure investment • Employment • Housing • Energy • Taxation	

Source: International Actuarial Association, Climate Risk Task Force (2020), *Importance of Climate-Related Risks for Actuaries*

Why actuaries should care?

- Reviewing the underlying models used in their work for their continued suitability in light of climate-related risks in the short and long terms – such a review may need to consider a system-wide approach to modelling climate-related risks
- Creating insurance products and pricing structures that align policyholders' financial interests with behaviour that promotes innovative solutions or climate-adaptive outcomes
- Aligning insurance product design with the needs of consumers, corporates, vulnerable groups, regulators, governments, etc.
- Encouraging pension funds, insurers and other clients to be active investors who support the management of climate-related risks in the companies in which they invest

Why actuaries should care? (cont’d)

- Sharing their expertise in modelling the financial impact of extreme climate-related events (e.g., catastrophe modelling)
- Developing investment strategies and products that will help address problems associated with climate-related risks
- Advising various types of organizations, including governments and other policymakers, on climate-related risk initiatives that encourage improved governance and risk management of this risk
- Contributing to the public debate and review of relevant government programs, public policy issues (e.g., supervision) and climate-related disaster planning, and informing building codes and land-use policies
- Disclosing in their work, in unambiguous terms, the impact that climate change has regarding the physical, transition and legal/reputation risks, according to frameworks (e.g., Task Force on Climate-Related Financial Disclosure [TCFD])

Substantiating the claim

- Groundbreaking analysis links over 5 million deaths annually to abnormally hot and cold temperatures
  - Source: Zhao, Q., Guo, Y., Ye, T., Gasparrini, A., Tong, S., Overcenco, A., … & Kinney, P. L. (2021). *Global, regional, and national burden of mortality associated with non-optimal ambient temperatures from 2000 to 2019: a three-stage modelling study*. The Lancet Planetary Health, 5(7), e415–e425
- Research on Hurricane Maria in Puerto Rico found 1,650 excess deaths in the six months following the storm, far exceeding the officially reported toll of 64
  - Source: Santos-Burgoa, C., Sandberg, J., Suárez, E., Goldman, A., Garcia-Meza, A., Pérez, C. M., ... & Zeger, S. L. (2018). *Differential and persistent risk of excess mortality from Hurricane Maria in Puerto Rico: A time-series analysis*. American Journal of Public Health, 108(9), 1202–1208
- Healthcare delivery disruption due to climate change
  - Source; Shah, A., Shapiro, J., & Hayes, S. (2022). *The Impact of Climate Change on Our Health and Health Systems*. The Commonwealth Fund
- Despite the rich empirical evidence, actuarial models still lack a practical, scalable framework to incorporate climate effects into pricing

Table 2
Death Incidence and Risk in the Six-Month Period Before (September 2016–March 2017) and After Hurricane Maria (September 2017–March 2018) by Cause of Death

Major Cause of Death Category	Pre-Maria		Post-Maria		Difference	
	Incidence	6-month risk per 100,000	Incidence	6-month risk per 100,000	Incidence Difference	Risk Ratio (95% CI)
CVD	2656	78.9	3023	92.7	367	1.18* (1.12, 1.24)
Alzheimers	1205	35.8	1437	44.1	232	1.23* (1.14, 1.33)
Diabetes Mellitus	1766	52.5	1966	60.3	200	1.15* (1.08, 1.23)
Sepsis	409	12.2	566	17.4	157	1.43* (1.26, 1.62)
Chronic respiratory disease	864	25.7	985	30.2	121	1.18* (1.08, 1.29)
Hypertension	710	21.1	811	24.9	101	1.18* (1.07, 1.31)
End Stage Renal Disease	490	14.6	556	17.1	66	1.17* (1.04, 1.32)
Suicide	104	3.1	142	4.4	38	1.41* (1.09, 1.82)
COPD	624	18.5	656	20.1	32	1.09 (0.98, 1.22)
Mental Health Conditions	282	8.4	311	9.5	29	1.14 (0.97, 1.34)
Asthma	60	1.8	58	1.8	-2	1.00 (0.70, 1.43)
Cancer	2883	85.7	2757	84.6	-126	0.99 (0.94, 1.04)
ALL CAUSES	16860	500.9	18510	567.8	1650	

*Significant RRs comparing post-Maria incidence risk with pre-Maria incidence risk

Source: Santos-Burgoa, C., Sandberg, J., Suárez, E., Goldman, A., Garcia-Meza, A., Pérez, C. M., ... & Zeger, S. L. (2018). *Differential and persistent risk of excess mortality from Hurricane Maria in Puerto Rico: A time-series analysis*. American Journal of Public Health, 108(9), 1202–1208

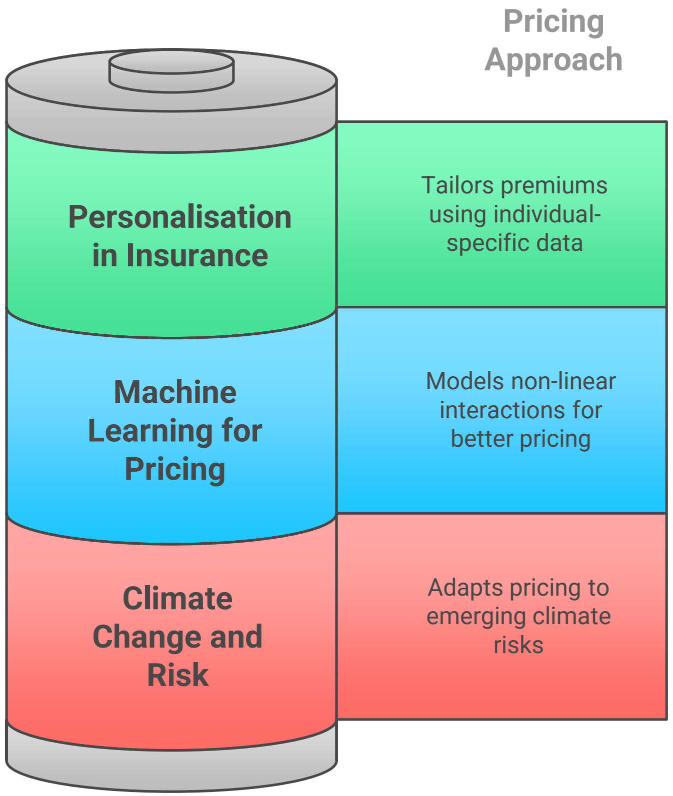
Types of mortality risk

Mortality risk is the risk that a portfolio will suffer from mortality being heavier than expected

Level	underlying mortality for a particular population dffers from that assumed
Volatility	mortality experience will differ from that assumed due to there being a finite number of lives in the population considered
Catastrophe	extreme version of volatility risk, i.e., risk of large losses due to some significant event increasing mortality rates beyond simple random volatility
Trend	mortality rates will change over time at a rate different to that assumend

Source: International Actuarial Association (2004). *A Global Framework for Insurer Solvency Assessment: Report of the Insurer Solvency Assessment Working Party*

Understanding insurance pricing:
From broad averages to personalisation



Data Overview

Personal data

Data theme	Example raw fields	Why GBMs handle it well
Demographics & family history	Age, gender (legal & identified), postcode-level socioeconomic indices, parental longevity	Mix of categorical + ordinal; GBMs capture non-linear age effects, regional interactions
Medical records & labs	ICD-10 diagnoses, prescription history, cholesterol, HbA1c, blood pressure	Lots of sparse codes + continuous labs; tree splits isolate rare but high-risk combos
Wearable / lifestyle	Daily step count, resting heart-rate trends, sleep efficiency, nicotine sensor	Highly granular, seasonally patterned; GBMs aggregate without formal time-series modelling
Behavioural / credit-like	Driving-telematics score, payment punctuality, online questionnaire honesty flag	Irregular, skewed; GBMs robust to outliers
Genomic polygenic risk scores *(opt-in)*	CAD-GRS, cancer-GRS percentiles	Weak marginal but strong interaction effects—GBMs pick these up

- All data must pass privacy, consent and anti-discrimination rules
- Public-domain data sources include WHO Mortality Database, CDC WONDER, etc.

Climate-Risk variables

Data theme	Example raw fields	How GBMs exploit them
Extreme-heat exposure	• Annual count of "wet-bulb > 26 °C days" in applicant's home 1 km grid (ERA5 / Copernicus) • 5-yr trend in summer night-time minima	Non-linear splits capture threshold effects ("mortality jumps once > 10 very hot nights/yr")
Wildfire smoke & air-quality	• Mean $PM_{2.5}$ concentration last 12 months • % of days with satellite-detected smoke plume (NOAA HMS)	Combines continuous (μg/m^3) & sparse binary plume flags; GBM finds their joint impact with age/COPD
Flood & storm surge risk	• Parcel-level FEMA Flood-Factor score • Count of *mandatory evacuations* within 5 km in past decade	Spatial categorical features interact with socio-economic indices (e.g., high-risk + low car ownership)
Vector-borne disease climate suitability	• Annual "days suitable for *Aedes aegypti*" • Tick-borne disease suitability index	GBM handles regional seasonality without explicit time-series modelling
Heat-related occupational exposure (if employer data available)	• % of work hours logged outdoors in > 30 °C • Presence of employer cooling-plan flag	Tree splits isolate high-risk occupation × climate interactions
Forward-looking scenario deltas	• Δ in 20-yr projected extreme-heat days under SSP2-4.5 (CMIP6 downscaling) • Δ wildfire probability 2030 vs 2020	Feeding scenario deltas lets GBM learn "future exposure" weightings alongside current

- Public-domain data sources include ERAS, NOAA NCEI, NASA POWER, local meteorological stations, etc.

ICD-10 codes that may be associated with climate variables

- Cardiovascular and Respiratory Diseases (ICD-10: I00–I99, J00–J99)
  - Heatwaves can increase mortality from cardiovascular events and respiratory distress, especially in vulnerable populations
  - Cold weather is also linked to increased deaths from heart attacks, strokes, and respiratory infections
  - Air pollution can exacerbate asthma, COPD and heart disease
- Infectious and Parasitic Diseases (ICD-10: A00–B99)
  - Climate variables like temperature, humidity, and precipitation influence the spread of vector-borne diseases (e.g., malaria)
  - Waterborne diseases can surge following heavy rainfall or flooding events
- External Causes of Mortality (ICD-10: V01–Y98)
  - floods, storms, heatwaves, wildfires
  - Seasonal patterns may influence accidents (e.g., icy roads in winter increasing vehicle-related deaths)
- Mental and Behavioural Disorders (ICD-10: F00–F99)
  - Some studies link seasonal affective disorder and suicide rates to temperature and sunlight exposure
- Neoplasms (ICD-10: C00–D48)
  - While less direct, UV radiation (linked to climate and geography) is a known risk factor for skin cancers

Toy data for demonstration purposes

```
df = pd.DataFrame(
    {
        # personal attributes
        'id': np.arange(1, N + 1),
        'age': age,
        'sex': sex,
        'region': regions,
        'bmi': bmi.round(1),
        'hypertension': hypertension,
        'diabetes': diabetes,
        'smoker': smoker,
        'avg_daily_steps': steps,
        'resting_hr': resting_hr,

        # climate-related features
        'heat_days': heat_days,
        'smoke_days': smoke_days,
        'chronic_pm25': chronic_pm25.round(2),
        'flood_risk_score': flood_risk_score.round(1),
        'wildfire_risk_score': wildfire_risk_score.round(1),
        'vector_disease_days': vector_disease_days,
        'projected_heat_delta': projected_heat_delta,

        # ICD-10 indicators
        'dx_T670': dx_T670,
        'dx_T675': dx_T675,
        'dx_T671': dx_T671,
        'dx_T68': dx_T68,
        'dx_T692': dx_T692,
        'dx_T691': dx_T691,

        # target
        'death_within_5yrs': death_within_5yrs
    }
)
```

- personal attributes
  - basic data → random deviates from uniform distributions
  - hypertension, diabetes, smoker probabilities → logistic functions and simulated binomial probabilities
  - daily steps, resting hours → random deviates from a Gaussian distribution, with means as functions of age, smoking indicator, hypertension
- climate-related features
  - heat days, smoke days → random deviates from Poisson distributions
  - chronic pm25, flood, wildfire, vector disease → random deviates from Gaussian distributions
  - projected heat delta → random deviates from uniform distribution
- ICD-10 indicators
  - arbitrary factors of heat days and cold days, combined with Bernoulli distributions to generate diagnosis indicators
- logit death
  - linear combination of all factors, with arbitrary coefficients
  - two-way interaction with high blood pressure and heat days
  - light label-flip noise to boost non-linearity
  - light miscoding noise for each ICD-10 toi account for false diagnoses
  - modest Gaussian noise for continuous variables

Sample use of logistic functions: hypertension & diabetes

```
hypertension_prob = expit(-7 + 0.06 * (age - 40) + 0.08 * (bmi - 25))
hypertension = np.random.binomial(1, hypertension_prob)
```

- Calibrated so as to
  - produce a low baseline probability at $\text{age} = 40$ and $\text{BMI} = 25$
  - adds $\sim 6\%$ to the log-odds per decade over 40, reflecting that hypertension accelerates with age
  - adds $\sim 8\%$ to the log-odds per $5\ \text{BMI}$ units above 25

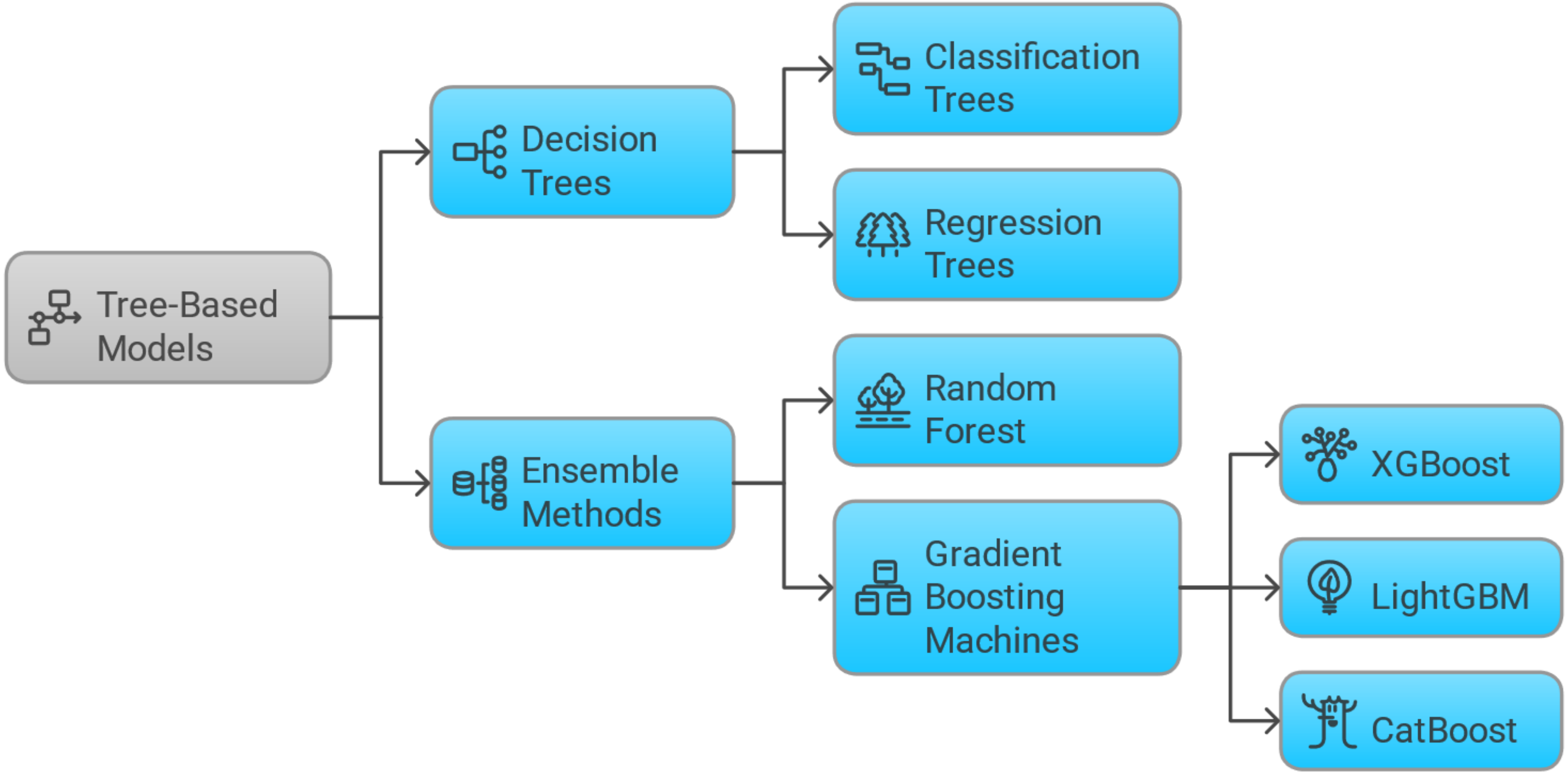
```
diabetes_prob = expit(-8 + 0.05 * (bmi - 28) + 0.04 * (age - 45))
diabetes = np.random.binomial(1, diabetes_prob)
```

- Calibrated so as to
  - recognise than diabetes is rarer than hypertension at baseline; thus, lower intercept
  - acknowledge that BMI is the dominant variable; $+5\ \text{BMI}$ lifts log-odds by 25%

Methodology

Tree-Based Models in Machine Learning

Selected models



Why use GBMs?

- Tabular data with complex relationships
  - GBMs are suitable on structured datasets with mixed data types and non-linear interactions
- Moderate-sized datasets
  - GBMs are data-efficient and work well with thousands of rows
- Handling of missing data and categorical values
  - **LightGBM** and **XGBoost** can handle missing values natively
  - **CatBoost** handles categorical data without encoding
- GBMs frequently win data science competitions (e.g., Kaggle) due to their flexibility and ensemble nature
- GBMs require minimum Python coding

From to R to Python

```r
# AdaBoost algorithm
T <- 50
models <- list()
alphas <- numeric(T)
weights <- rep(1/nrow(train), nrow(train))

for (t in 1:T) {
  stump <- build_stump(train, train$y, weights)
  error <- stump$error

  if (error == 0 || error > 0.5) break

  alpha <- 0.5 * log((1 - error) / error)
  alphas[t] <- alpha
  models[[t]] <- stump

  # Update weights
  weights <- weights * exp(-alpha * train$y * stump$pred)
  weights <- weights / sum(weights)
}

# Prediction function
predict_boost <- function(models, alphas, data) {
  final_pred <- rep(0, nrow(data))
  for (i in 1:length(models)) {
    m <- models[[i]]
    pred <- ifelse(m$direction * data[[m$feature]] < m$direction * m$threshold, 1, -1)
    final_pred <- final_pred + alphas[i] * pred
  }
  return(ifelse(final_pred >= 0, 1, -1))
}

# Predict and evaluate
test$pred <- predict_boost(models, alphas, test)
accuracy <- mean(test$pred == test$y)

# Create decision grid
grid <- expand.grid(
  x1 = seq(min(df$x1), max(df$x1), length.out=200),
  x2 = seq(min(df$x2), max(df$x2), length.out=200)
)
```

```python
from xgboost import XGBClassifier

classifier = XGBClassifier(
    n_estimators=100,
    learning_rate=0.1,
    random_state=42,
    eval_metric='logloss'
    )

classifier.fit(X_train, y_train)
```

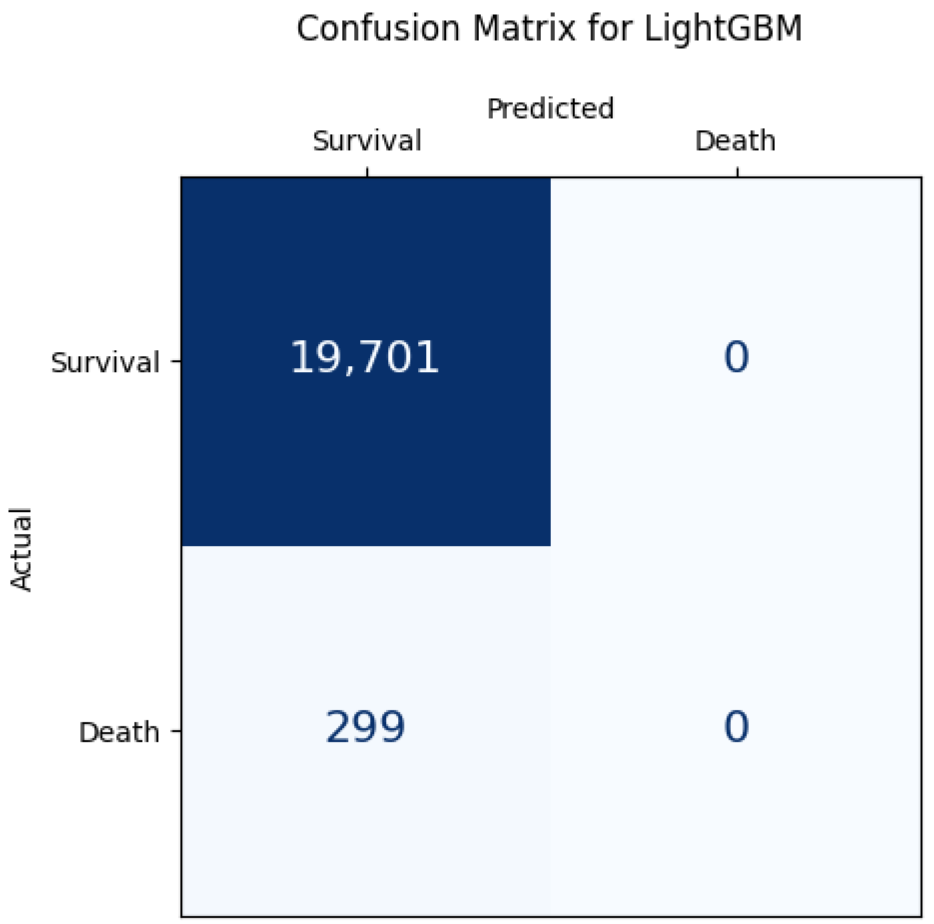
Caveat Emptor

```
class sklearn.ensemble.GradientBoostingClassifier(*, loss='log_loss',
learning_rate=0.1, n_estimators=100, subsample=1.0, criterion='friedman_mse',
min_samples_split=2, min_samples_leaf=1, min_weight_fraction_leaf=0.0, max_depth=3,
min_impurity_decrease=0.0, init=None, random_state=None, max_features=None,
verbose=0, max_leaf_nodes=None, warm_start=False, validation_fraction=0.1,
n_iter_no_change=None, tol=0.0001, ccp_alpha=0.0)
```

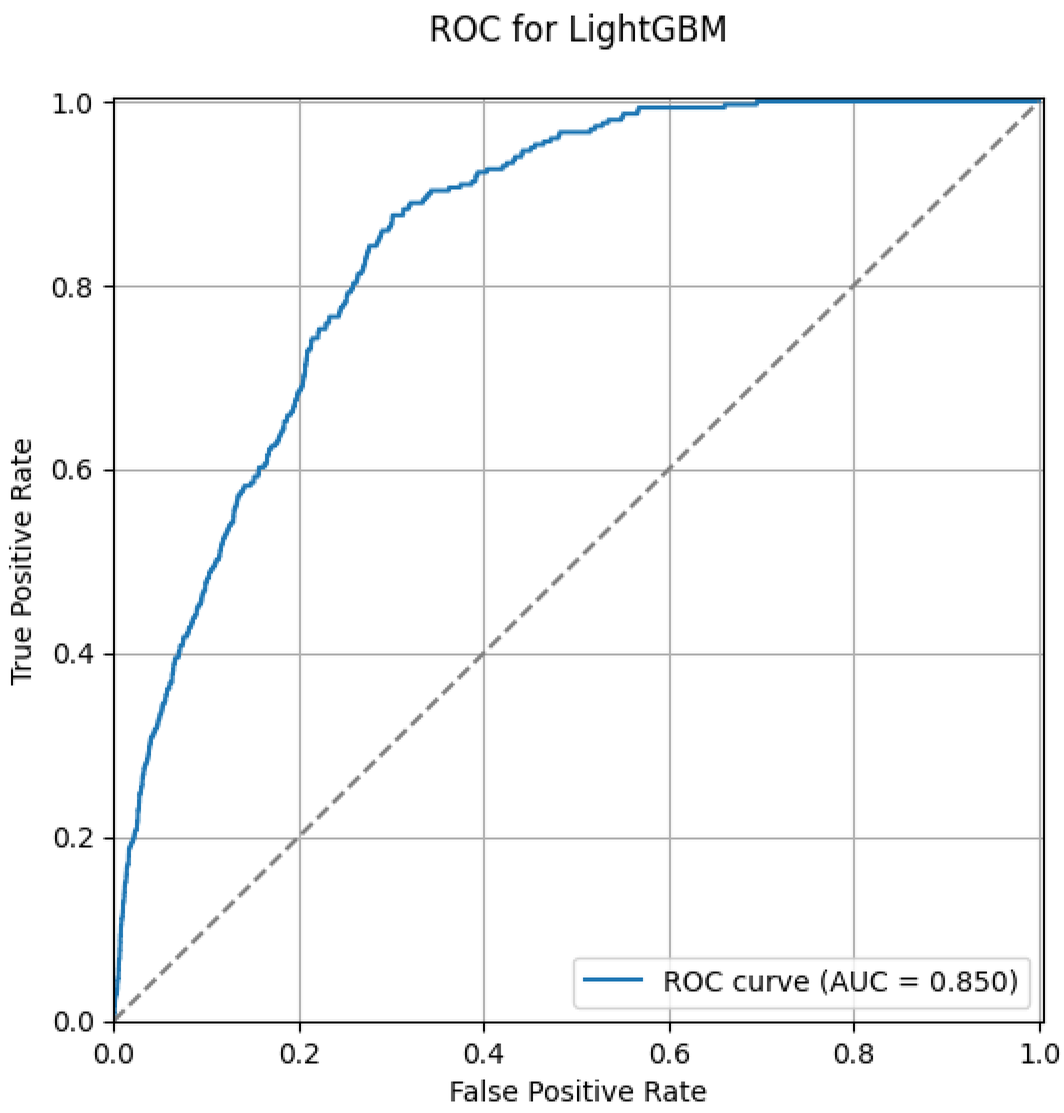
[source]

- The challenge has shifted from coding to understanding the impact of the parameters in the output!
- Fortunately, `scikit-learn.org` provides fairly detailed documentation
  - ditto for other libraries, such as `xgboost` and `lightgbm`

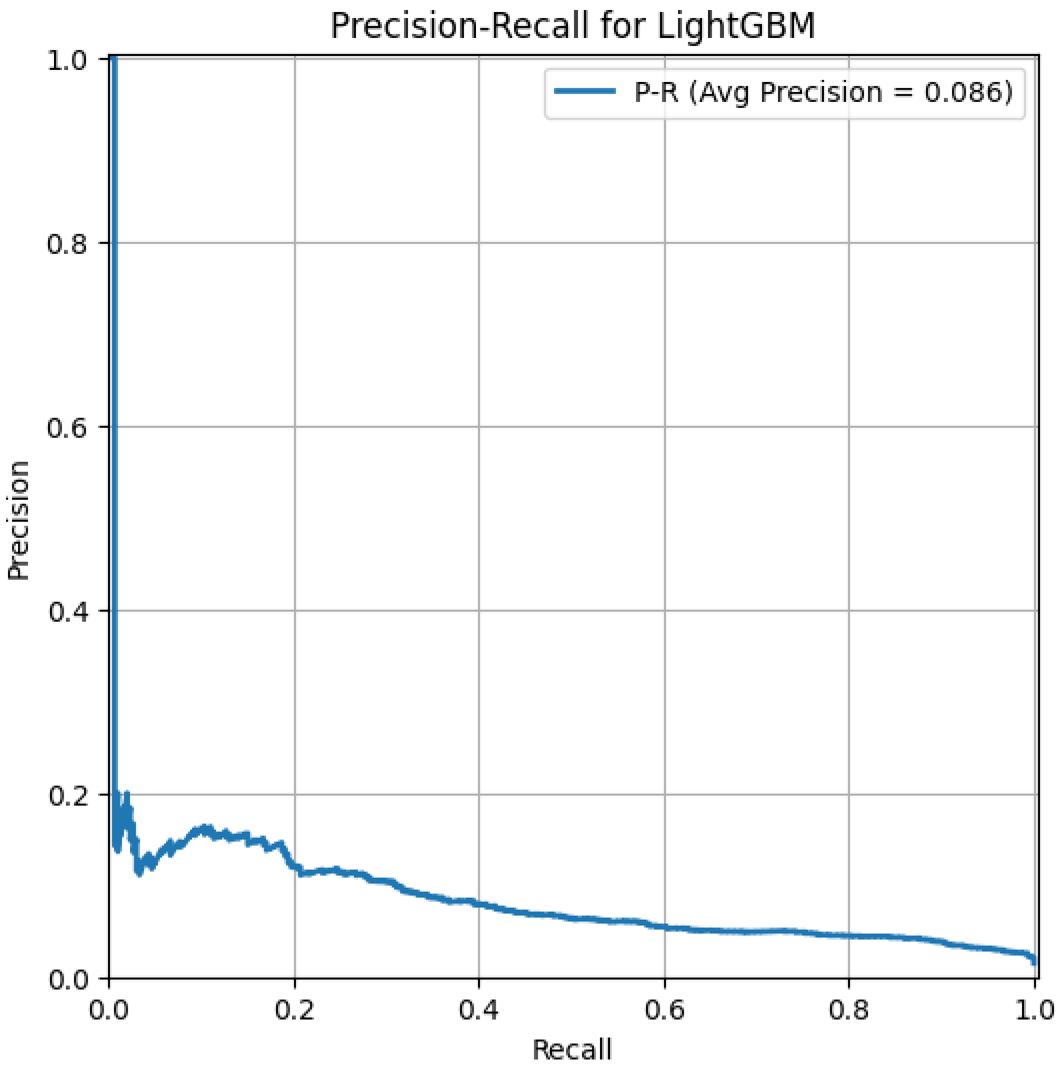
Results



1. Diagonal cells: correct classifications
2. Off-diagonal cells: misclassifications
3. Perfect model: identity matrix scaled by class counts



1. A perfect model would hug the top-left corner of the plot and have an AUC of 1
2. The dashed gray line represents random guessing, with AUC of 0.5



1. A perfect model would have a P-R curve that goes straight across the top and then straight down

2. Avg Precision is best for imbalanced datasets

Metrics

model	accuracy	precision	recall	f_1	roc_auc	pr_auc
Gradient Boosting	0.985	0.970	0.985	0.978	0.849	0.083
XGBoost	0.747	0.981	0.747	0.841	0.844	0.080
LightGBM	0.985	0.970	0.985	0.978	0.850	0.086
CatBoost	0.799	0.979	0.799	0.875	0.778	0.064

Accuracy: % of total predictions that were correct

Precision: When the model says "positive," how often is it right?

Recall: Of all actual positives, how many did the model find?

F1 Score: Weighted average between precision and recall

ROC AUC: Ability to distinguish between classes (higher = better)

PR AUC: Quality of model when dealing with imbalanced data (higher = better)

How do models stack up?

Gradient Boosting	Accuracy ($\mathbf{0.985}$): Almost perfect Precision & Recall ($\approx \mathbf{0.97 - 0.99}$): High ability to correctly detect positives F1 ($\mathbf{0.978}$): Balanced performance ROC AUC ($\mathbf{0.849}$): Good class discrimination PR AUC ($\mathbf{0.083}$): Very low, raises red flags for imbalance
XGBoost	Accuracy ($\mathbf{0.747}$): Noticeably lower Precision ($\mathbf{0.981}$): Very confident in its positives Recall ($\mathbf{0.747}$): Misses many true positives F1 ($\mathbf{0.841}$): Decent, but hurt by low recall ROC AUC ($\mathbf{0.844}$): Acceptable PR AUC ($\mathbf{0.080}$): Very low, raises red flags for imbalance
LightGBM	Accuracy ($\mathbf{0.985}$): Almost perfect Precision & Recall ($\approx \mathbf{0.97 - 0.99}$): Matches Gradient Boosting F1 ($\mathbf{0.978}$): Top-tier ROC AUC ($\mathbf{0.850}$): Slightly best PR AUC ($\mathbf{0.086}$): Still very low, but slightly better than others
CatBoost	Accuracy ($\mathbf{0.799}$): Better than XGBoost but well below GBM/LightGBM Precision ($\mathbf{0.979}$): High confidence Recall ($\mathbf{0.799}$): Misses positives F1 ($\mathbf{0.875}$): Acceptable ROC AUC ($\mathbf{0.778}$): Lower class separation PR AUC ($\mathbf{0.064}$): Lowest, again points to class imbalance

Plus emptoris caveat

- Class imbalance
  - makes accuracy misleading (if 99% are survivals, predicting survival gives 99% accuracy but zero usefulness)
  - all models have a low PR AUC score, which is normal due to a highly imbalanced data set
  - may resample (e.g., Synthetic Minority Over-Sampling Technique [SMOTE]) or adjust thresholds
- Interpretability
  - GBMs are not very transparent
  - may use Shapley Additive Explanations (SHAP) to learn "why did the model make this prediction"
  - SHAP explains prediction clearly and fairly, and work especially well with GBMs—explainability is critical in insurance!
- Computational resources
  - LightGBM is optimised for large data sets, making it more efficient than other GBM algorithms
  - if the model is to be used in real-time or resource-limited environments, LightGBM is lightweight and deployable
  - scikit-learn GBMs are not optimised for speed, early stopping must be manually implemented, and there is no handling of missing values (pre-imputation must be done by the user)
- Model tuning
  - Hyperparameters should be fine-tuned with cross-validation

May go for LightGBM as a starting model, but ...

- Address class imbalance
- Use cross-validation
  - split data set into k-sized folds
  - use each fold as the validation set, and the remaining $k-1$ folds as the training set
  - train and evaluate the model k times, and average the performance across all folds
- Fine-tune parameters
  - adjust settings that control how the model learns, in order to improve its performance (e.g., no. of boosting rounds [trees], how much each tree corrects the previous one, maximum depth of each tree, etc.)
- Use explainability tools, like SHAP

Common pain points and potential fixes

Pain point	Reality check	Practical fix
Interpretability	Raw GBM trees aren't transparent to clinicians or regulators	SHAP value plots, partial-dependence curves, monotonic constraints, or a post-hoc logistic "proxy" model may satisfy scrutiny
Calibration	Boosters can be over-confident, especially with rare outcomes	scikit-learn calibration guide explains how to use built-in calibration layers (e.g., CalibratedClassifierCV)
Censoring & time horizon	A plain GBM treats mortality as a *binary* label; life actuaries care about *time to event*	Use survival GBMs: LightGBM's "objective": "survival", XGBoost's "survival:cox", scikit-survival's GradientBoostingSurvivalAnalysis, or a Poisson-GBM on person-time intervals
Regulatory fairness	Boosting can inadvertently encode socio-economic or racial bias	Constrain or exclude protected features; run fairness dashboards
"All models tie"	Occurs when data are too easy, class imbalance unaddressed, or metrics use a single threshold	Re-balance classes, tune hyper-parameters, pick model-specific thresholds, or use more challenging non-linear signal

Conclusions

Take-aways

- Extreme temperatures, storms, and chronic environmental stressors significantly impact mortality—often underreported and underestimated
- Mortality impacts from climate events often go unaccounted in standard actuarial models, necessitating richer, more granular data sources and scenario analysis
- Actuaries need decision-making frameworks that combine quantitative rigor with qualitative insights
- Both acute events (e.g., hurricanes) and long-term exposures (e.g., heatwaves, pollution) must be modeled to capture the full climate-health burden
- Embedding climate-health projections into actuarial models empowers insurers and policymakers to improve risk management, product design, and societal resilience
- By capturing complex, non-linear interactions between demographic, environmental, and exposure variables, GBMs help actuaries model mortality impacts more accurately and transparently