EKEYA.GLOBAL HOLDINGS

ENGINEERING · DATA ANALYTICS

CONSTRUCTION MATERIALS ANALYTICS

DATA ANALYTICS · CASE STUDY
REF. EKEYA / DA-2026-01

Predicting Concrete Compressive Strength

A machine-learning model that estimates 28-day strength directly from a mix design — turning a 28-day lab wait into a same-day decision for quality control, cost control, and mix optimization.

0.93 R^2	±3.0 MPa	1,030	8 → 13
VARIANCE EXPLAINED	TYPICAL ERROR (MAE)	LAB MIXES ANALYZED	FEATURES (DOMAIN-ENGINEERED)

Compressive strength is the number that decides whether a concrete mix is safe to pour — but confirming it means casting cylinders and crushing them after **28 days of curing**. That wait stalls schedules and locks in a mix chosen weeks earlier. This study builds a model that predicts strength from the mix proportions and age alone, accurately enough to guide decisions before the lab results return.

01 Why it matters

- **Quality control** — estimate strength on day 1 and flag a weak batch before it is poured, instead of discovering it 28 days later.
- **Cost & carbon** — cement is the most expensive and most carbon-intensive ingredient. The model lets you search for leaner mixes that still hit the target strength.
- **Mix design** — replace trial-and-error batching with data-driven optimization.

02 The data

1,030 laboratory concrete mixes (I-Cheng Yeh, UCI Machine Learning Repository, CC BY 4.0): cement, blast-furnace slag, fly ash, water, superplasticizer, coarse and fine aggregate, and curing age — with measured compressive strength in MPa. The data is complete, though it contained 25 duplicate mixes, which were removed to prevent identical rows from leaking across the train–test split. Strength ranges from lightweight fill to high-performance concrete.

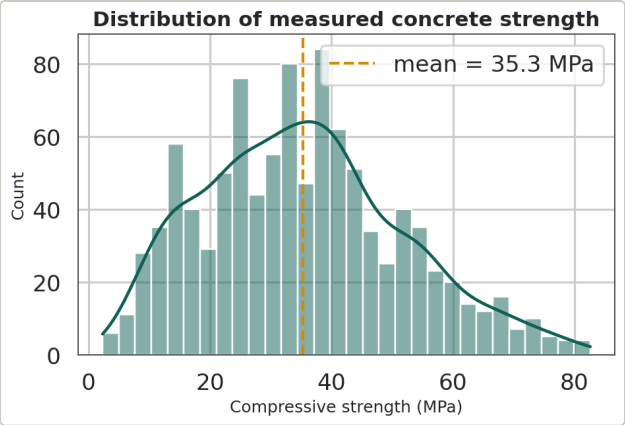

Fig 1 — Strength spans ~2 to 83 MPa.

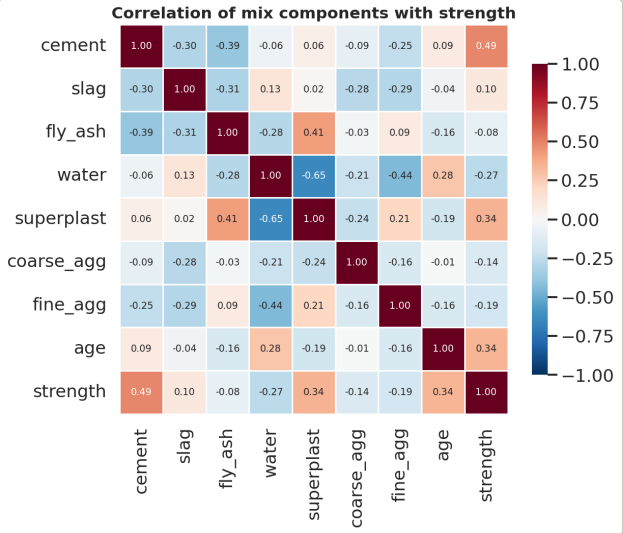

Fig 2 — Cement helps, water hurts; no single ingredient decides strength.

03 Materials knowledge, encoded as features

The decisive step is translating materials science into inputs a model can use. From the raw quantities we derive the **water/cement** and **water/binder** ratios (the classic and SCM-aware forms of Abrams' law), total **binder** content, and the **supplementary-materials fraction**. These are the levers an engineer actually turns — and, as Section 5 shows, they are exactly what the model relies on most.

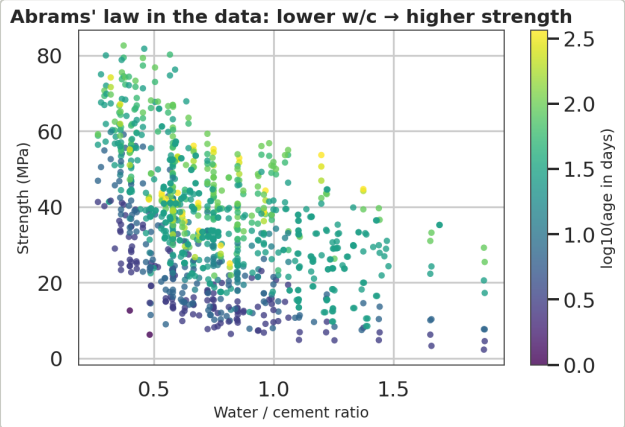

Fig 3 — Abrams' law is visible: lower water/cement → higher strength.

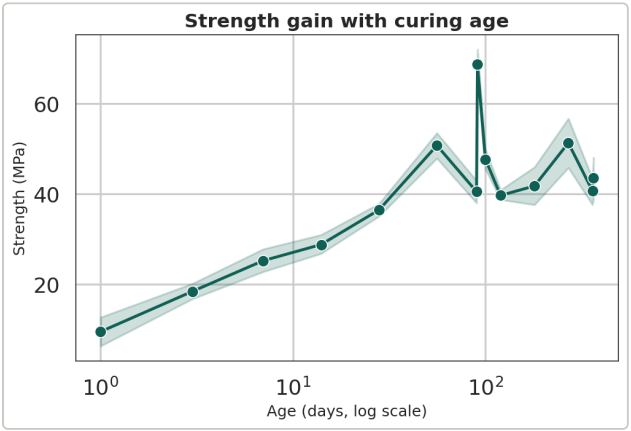

Fig 4 — Strength climbs with curing age, fast then leveling off.

04 Model performance

Six models were compared with 5-fold cross-validation and a held-out test set. Linear models plateau near **R² ≈ 0.59** — strength is too nonlinear for them. Tree ensembles capture the interactions and reach **R² ≈ 0.91–0.93**. The best, a histogram gradient-boosting model, predicts strength with a typical error of about **±3 MPa** across a 2–83 MPa range, with unbiased residuals.

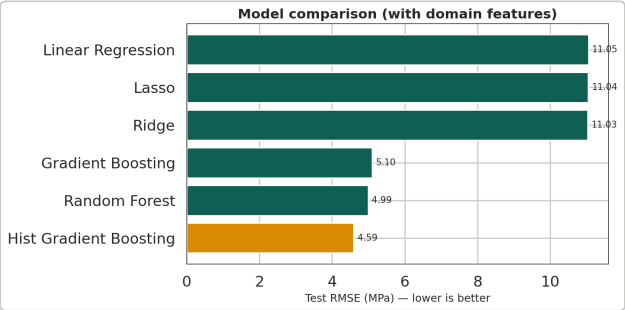

Fig 5 — Test RMSE by model (lower is better).

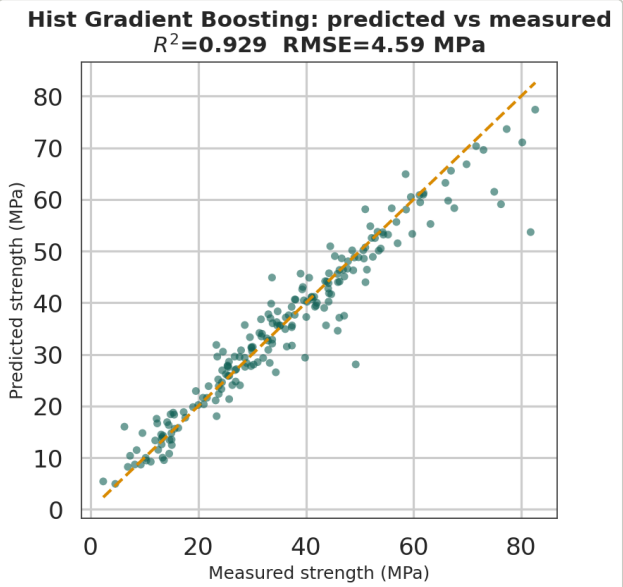

Fig 6 — Predictions track the measured values closely.

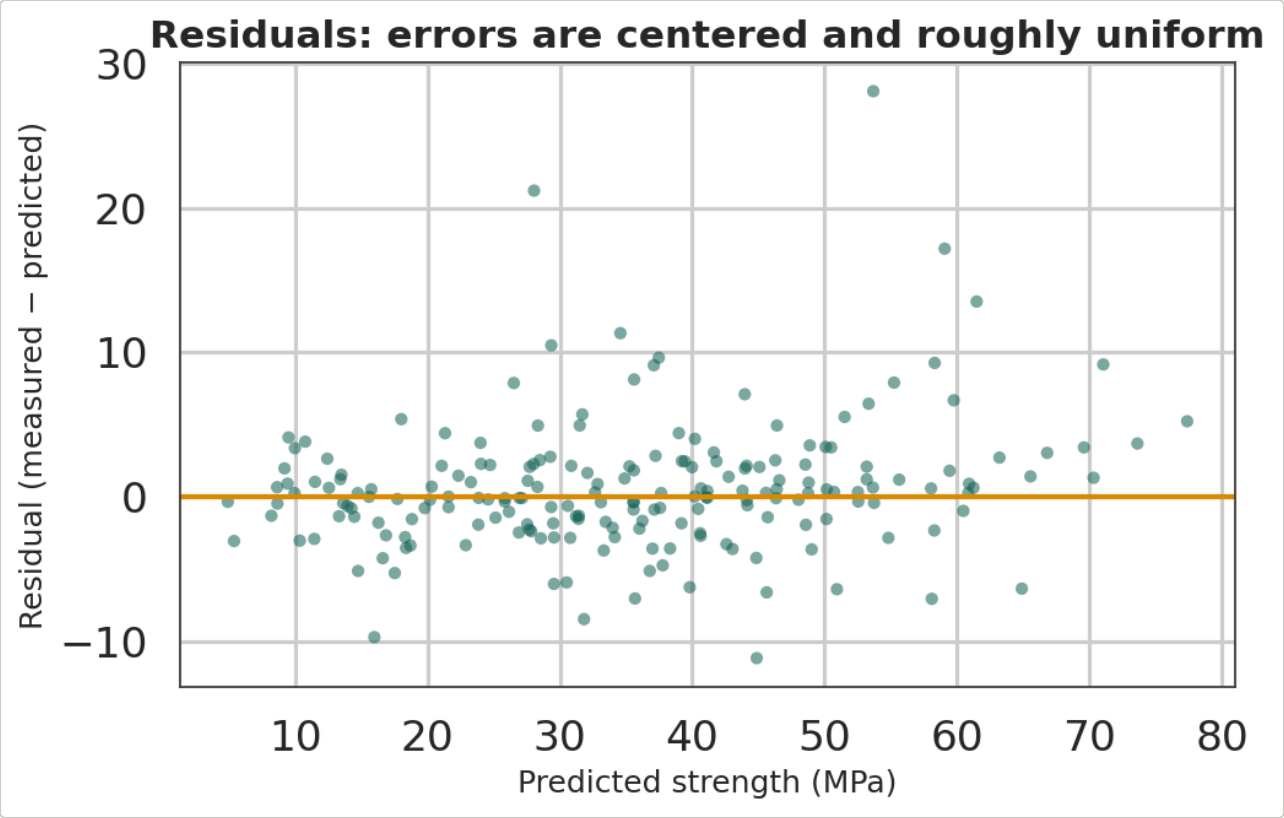

Fig 7 — Residuals are centered on zero with no systematic pattern — the model is reliable across the full range.

05 What drives strength

The model leans hardest on **curing age** and on the **water/binder** and **water/cement** ratios — two features that do not exist in the raw data and were engineered from domain knowledge. The physics that governs concrete is the same signal the algorithm finds most useful: materials expertise and machine learning point to the same answer.

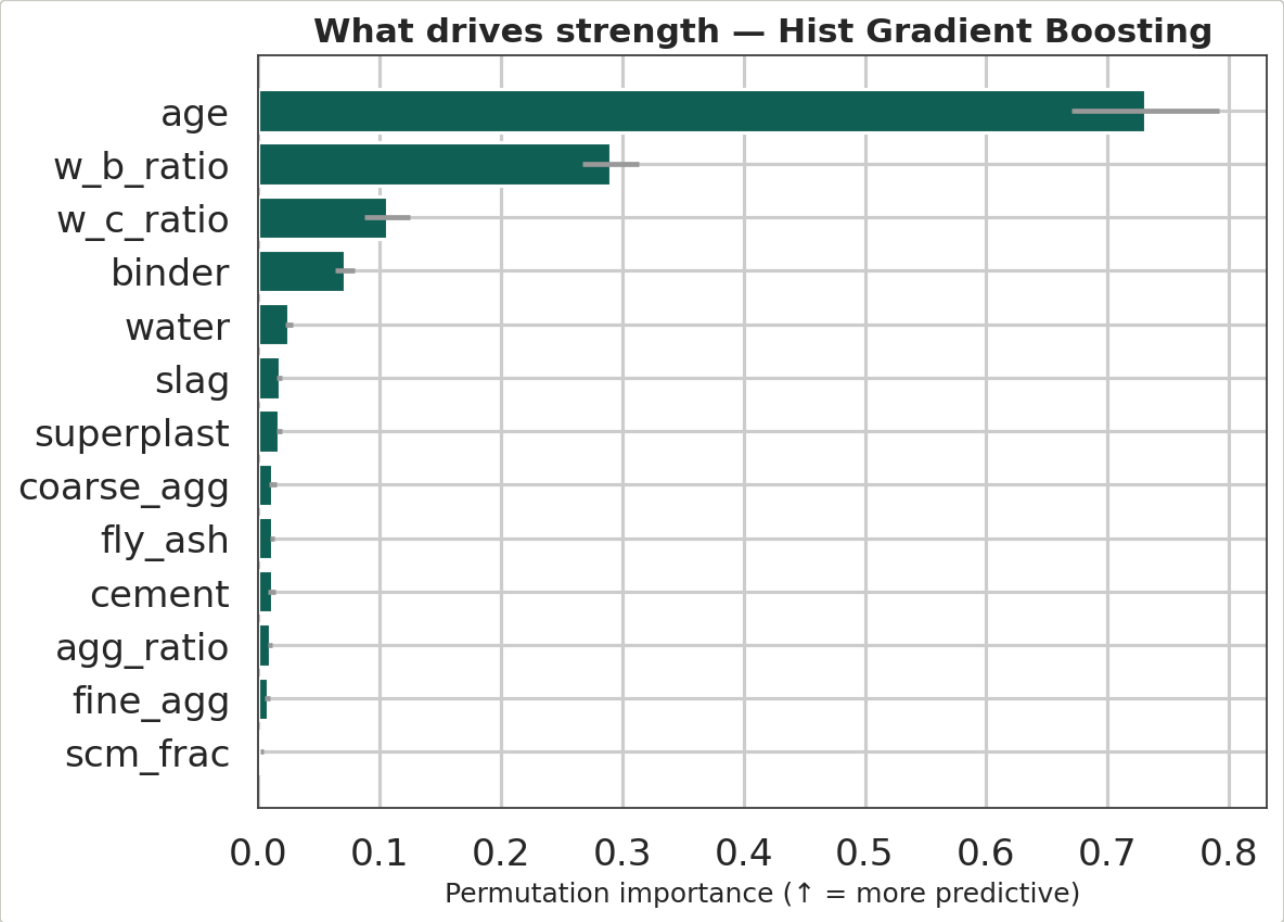

Fig 8 — Permutation importance. Domain-engineered ratios rank among the top predictors.

06 Business value & next steps

> **Bottom line:** from a mix design and an age, this model predicts 28-day compressive strength within about ±3 MPa — accurate enough to support same-day QC decisions, cement-reduction studies, and data-driven mix optimization.

- **Mix optimizer** — "the cheapest / lowest-carbon mix that still reaches 35 MPa at 28 days."
- **Per-prediction explanations** (SHAP) so each estimate comes with its reasoning.
- **A simple web tool** that returns predicted strength from any mix, for the lab or the plant.