## Binomial GLM – Email Open Rate Model

* **Goal:** Predict email open rates (OR) using 13 features
* **Model Type:** Binomial GLM with logit link — appropriate for proportion data like open rate.
* **Data Setup:** Cleaned, winsorized, and one-hot encoded with ~13 categorical and numeric predictors.
* **Accuracy:**
  + Mean Actual OR = **0.1275**
  + Mean Predicted OR = **0.1109**
  + **MAE = 0.0285**, **RMSE = 0.0380** (≈3–4% error range).
* **Calibration:** Predictions align well with actuals across most deciles; mild underestimation at the top end.
* **Top-Decile Lift:**
  + Top 10% predicted OR = **0.1498** vs overall 0.1153 → **1.3× lift** in open rate.
* **Insights:** Model effectively distinguishes higher-performing campaigns; personalization and tone likely key drivers.
* **Use Case:** Estimate expected open rate for new email variants; guide A/B testing and targeting.
* **Next Steps:** Add interactions or advanced models (e.g., LightGBM) to capture nonlinear effects.

## LightGBM Model – Email Open Rate Prediction Summary

* **Model Type:** Gradient Boosted Trees (LightGBM) – captures nonlinear feature effects and interactions automatically.
* **Accuracy:**
  + **MAE = 0.0193**, **RMSE = 0.0246**
  + Weighted MAE = **0.0128**, Weighted RMSE = **0.0169**
  + Explains about **43% of variance (R² = 0.43)** on hold-out data.
* **Calibration:** Predicted open rates closely align with actuals across deciles — well-calibrated even in high-performance bins.
* **Top-Decile Lift:**
  + **Unweighted Lift:** 1.42× (0.150 vs 0.106 overall)
  + **Weighted Lift:** 1.59× (0.151 vs 0.095 overall) — strong ability to rank top-performing campaigns.
* **Insights:**
  + Model effectively identifies high-engagement campaigns.
  + Handles nonlinearities and feature interactions better than GLM.
* **Business Impact:** Enables smarter targeting and prioritization of campaigns with the highest open rate potential.

## 💡 Key Insights

* **Predictive Power:**
  + LightGBM substantially improves prediction accuracy — about **30–40% lower error** than the GLM.
  + Explains **43% of open rate variation**, versus limited explanatory power in the GLM.
* **Ranking Performance:**
  + Both models successfully rank campaigns by expected performance.
  + LightGBM’s **1.59× top-decile lift** means the best-predicted 10% of campaigns achieve open rates **~60% higher** than average.
* **Calibration & Realism:**
  + GLM underestimates high-performing campaigns slightly.
  + LightGBM predictions closely mirror actual open rates across all deciles.
* **Interpretability vs. Power:**
  + **GLM:** Easy to interpret; identifies how each variable (e.g., personalization, tone, emoji) affects open rate directionally.
  + **LightGBM:** More accurate and flexible but less directly interpretable; feature importance analysis is recommended.

## Business Implications

* Both models enable **data-driven optimization** of campaign design and timing.
* LightGBM can be deployed for **automated scoring** of new campaigns — predicting expected open rate before launch.
* GLM remains valuable for **insight generation** and **explainability** — ideal for strategy and communication teams.
* Combining both:
  + Use **LightGBM for prediction**,
  + Use **GLM for explanation and validation**.