Who Says “Yes”?
A Machine Learning Approach to Bank Marketing

Presenter: Zhang Jingyi

Core XGBoost Objective

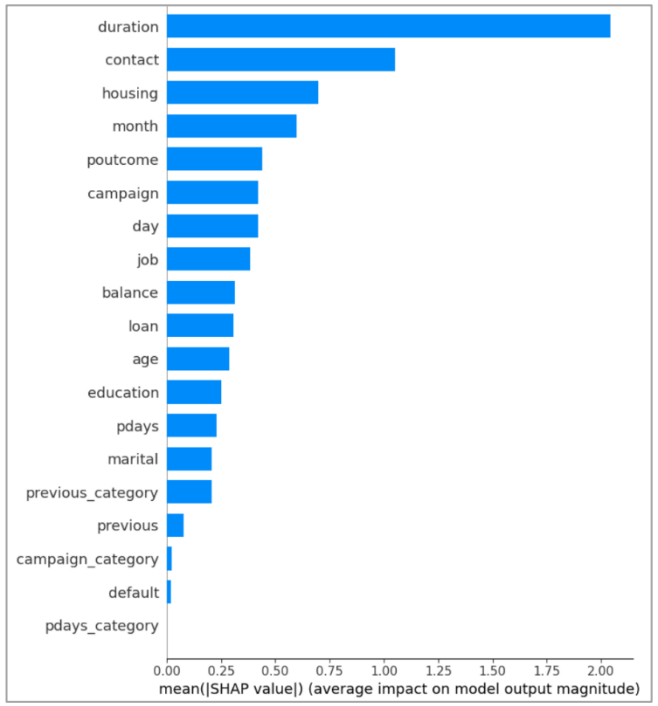
$$L^{(t)} \approx \sum_{i=1}^{n} \left[g_i f_t(x_i) + \frac{1}{2} h_i f_t(x_i)^2 \right] + \Omega(f_t)$$

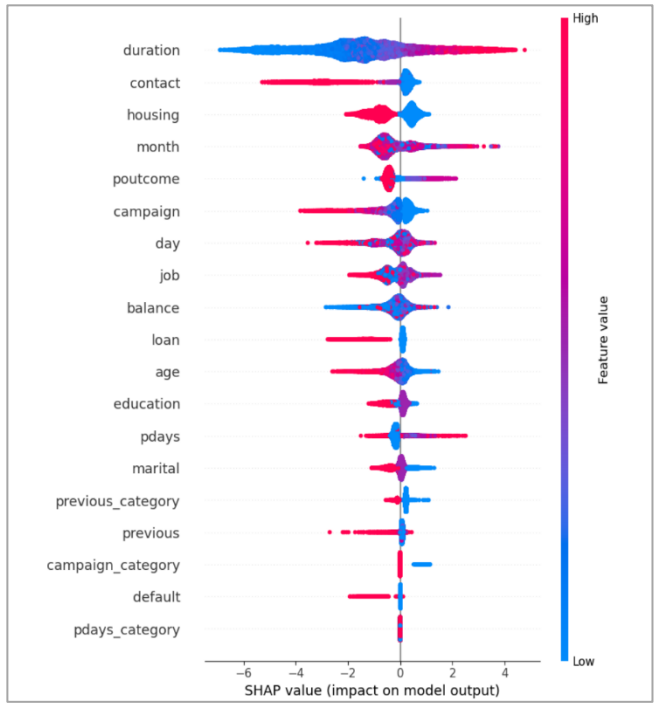
- g_i: First-order gradient (direction)
- h_i: Second-order gradient (step size)
- $\Omega(f_t)$: Regularization (prevent overfitting)

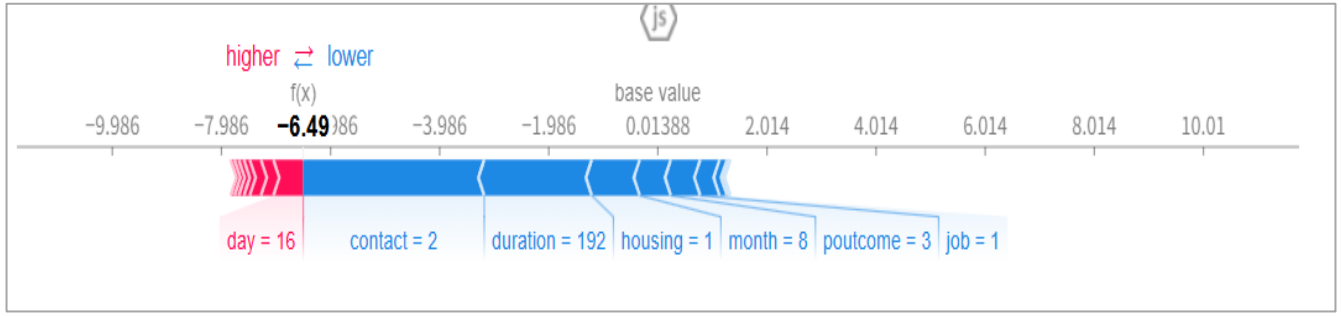
- **Demystifying SHAP**
- **Result Analysis:**

Model	MCC	F1-score	Precision	Recall	Accuracy
XGBoost (Full features)	0.5151	0.5961	0.5076	0.662	0.8537
XGBoost (Select Features)					
XGB+RF					

SHAP Analysis: The Key to Opening the ML Black Box

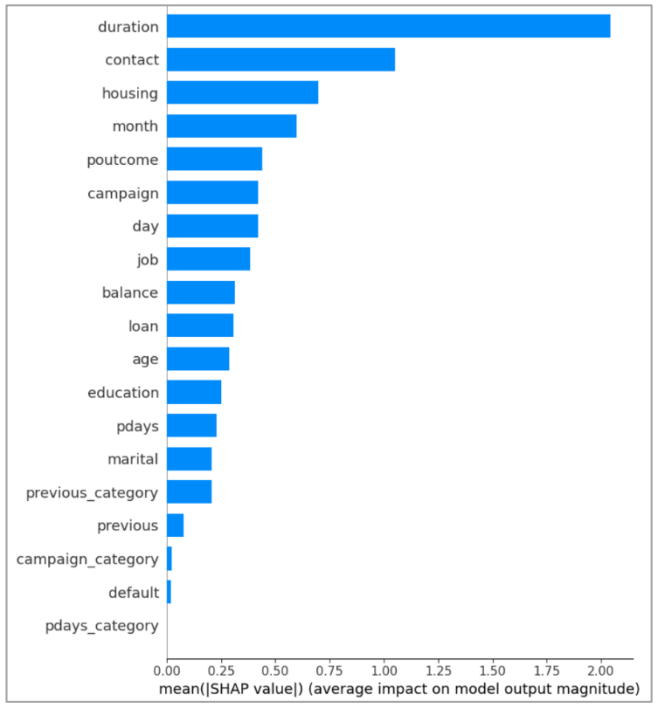


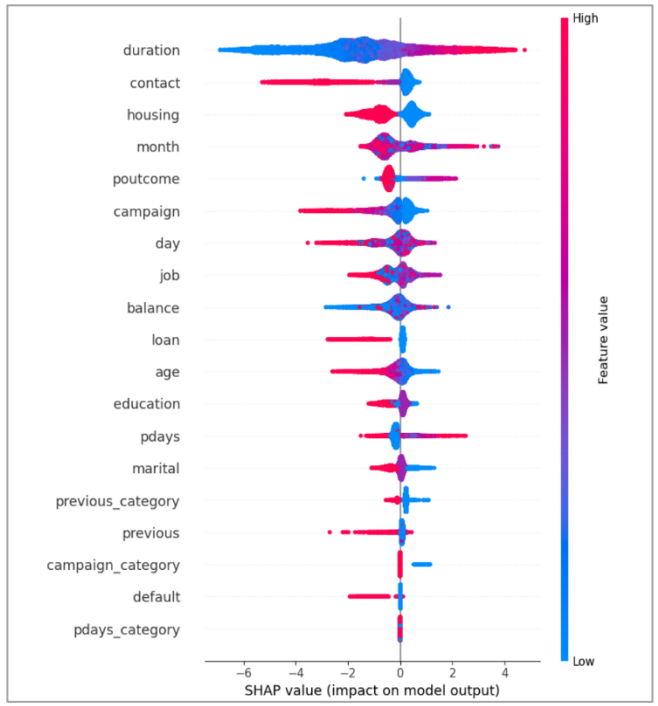


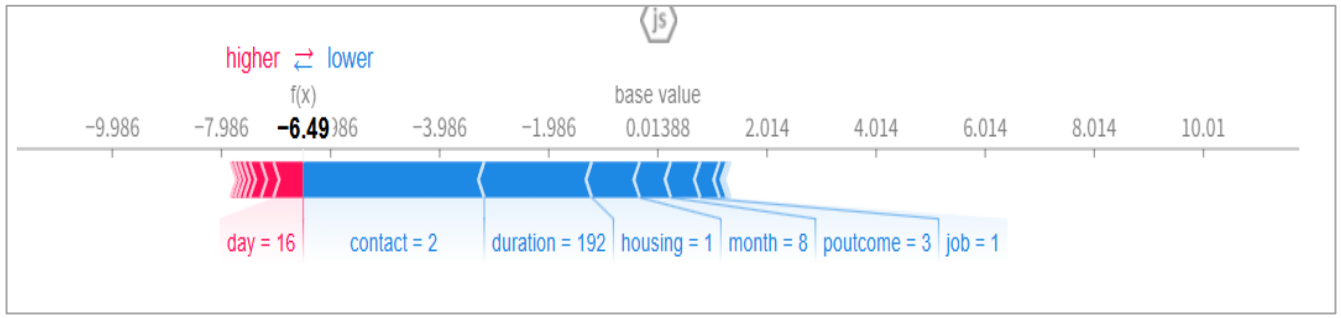


- **Key Driver**
  - Duration
  - Engaged customers talk longer.
- **Impact Direction:**
  - Sales should focus on retaining customers with long call times.
- **Individual Instance Explanation:**
  - A call lasting **192 seconds**.
  - decreased the predicted likelihood of subscription.

SHAP Analysis: The Key to Opening the ML Black Box







SELECTED FEATURES

duration, housing, poutcome, month, contact, previous, campaign, age, education

Core XGBoost Objective

$$L^{(t)} \approx \sum_{i=1}^{n} \left[g_i f_t(x_i) + \frac{1}{2} h_i f_t(x_i)^2 \right] + \Omega(f_t)$$

- g_i: First-order gradient (direction)
- h_i: Second-order gradient (step size)
- $\Omega(f_t)$: Regularization (prevent overfitting)

✓ Demystifying SHAP

—From Black Box to Transparency

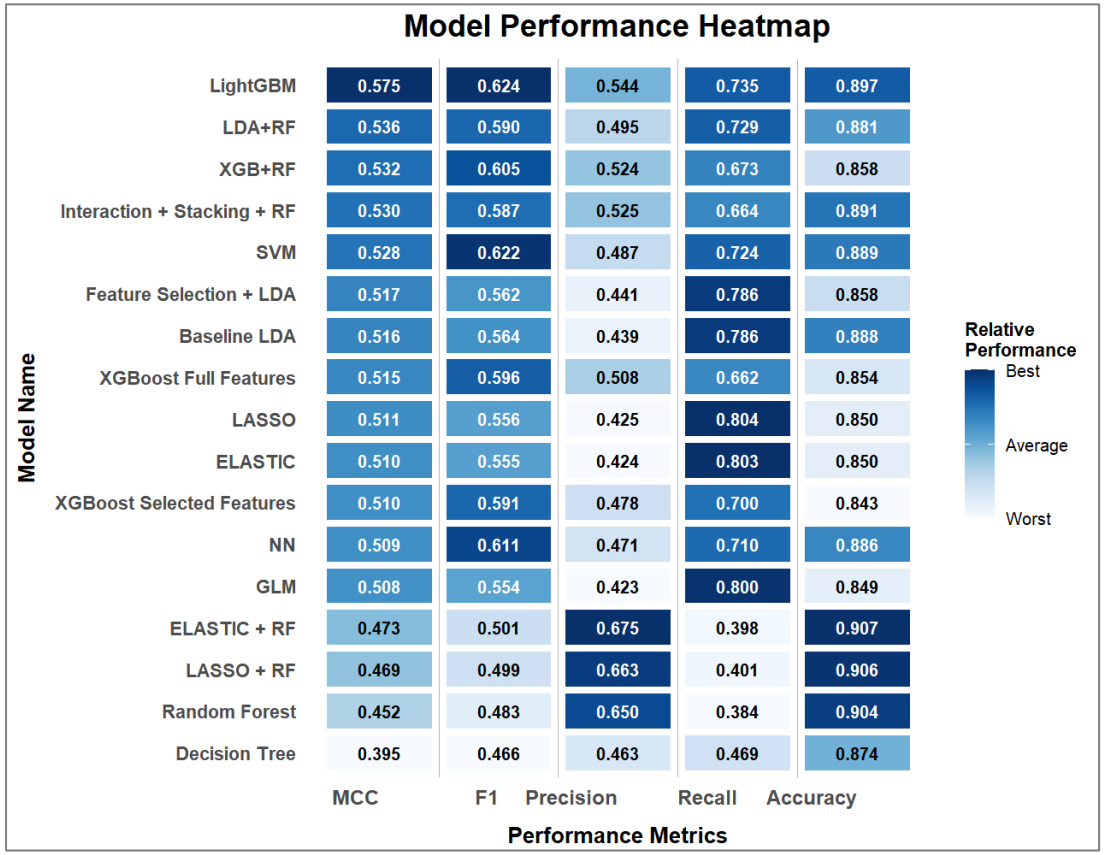
✓ Result analysis

—**Full Features:** Baseline with noise.

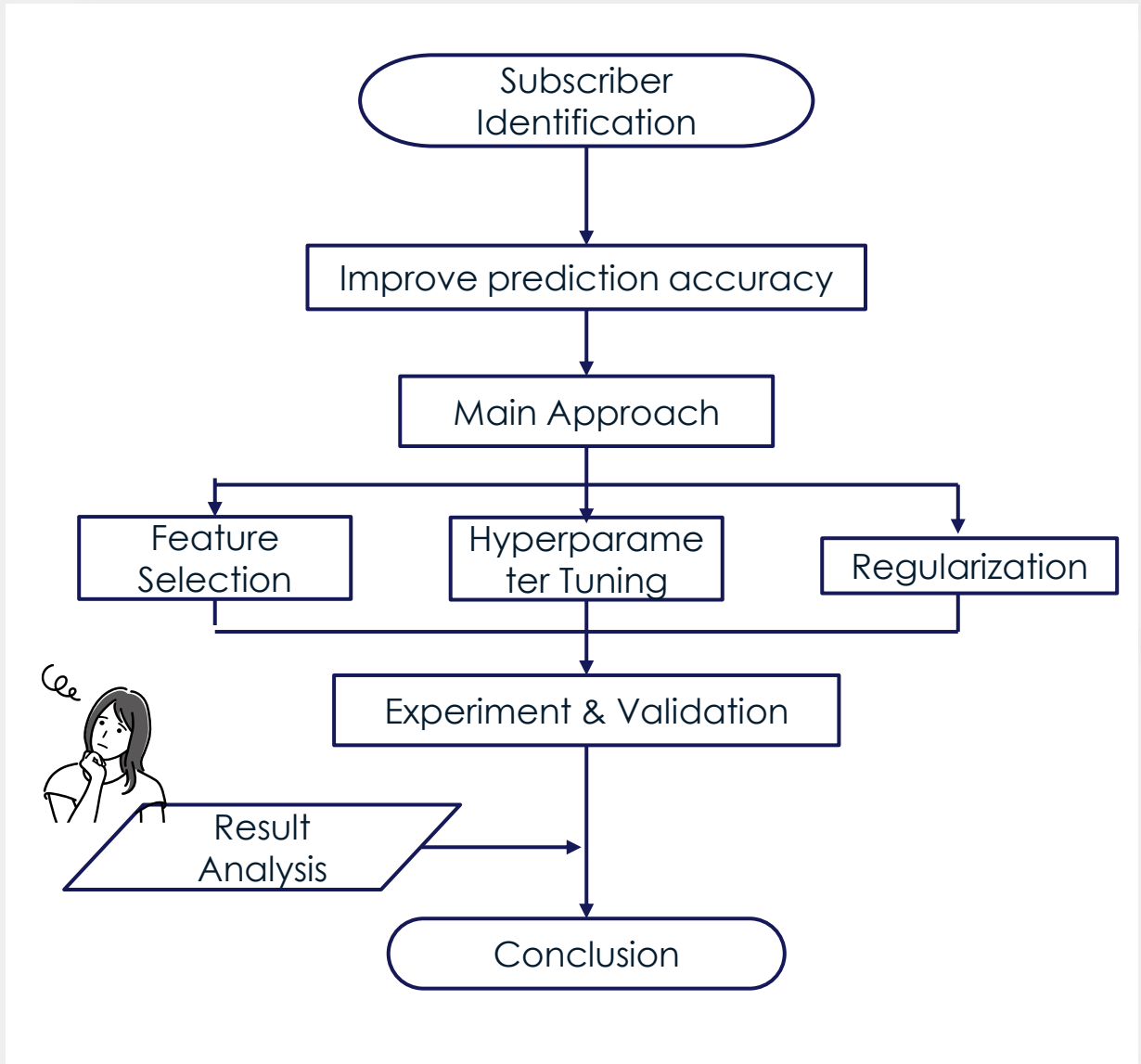
—**Feature-Selected:** ↑Recall, ↓Precision.

—**Ensemble (Optimal):** ↑Overall Performance

Model	MCC	F1-score	Precision	Recall	Accuracy
XGBoost (Full features)	0.5151	0.5961	0.5076	0.662	0.8537
XGBoost (Select Features)	0.5096	0.5908	0.4776	0.7001	0.8432
XGB+RF	0.532	0.6045	0.5242	0.6728	0.8575



MCC ≈ 0.5: A Sign of Real-World Data



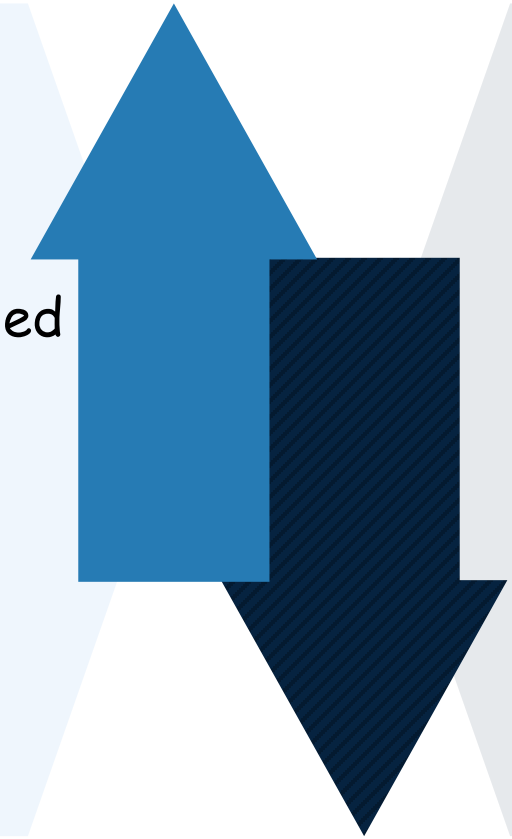
NANYANG TECHNOLOGICAL UNIVERSITY SINGAPORE

The Business Trade-off: Precision vs. Recall

Precision

"How many of the users we called were actually subscribers?"

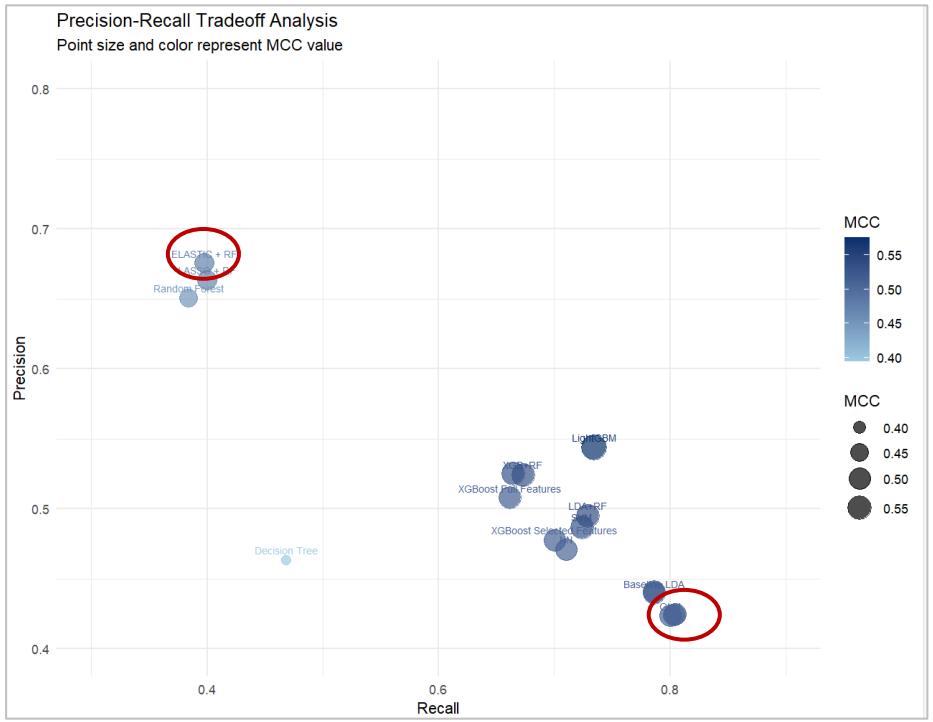
It's about cost efficiency.



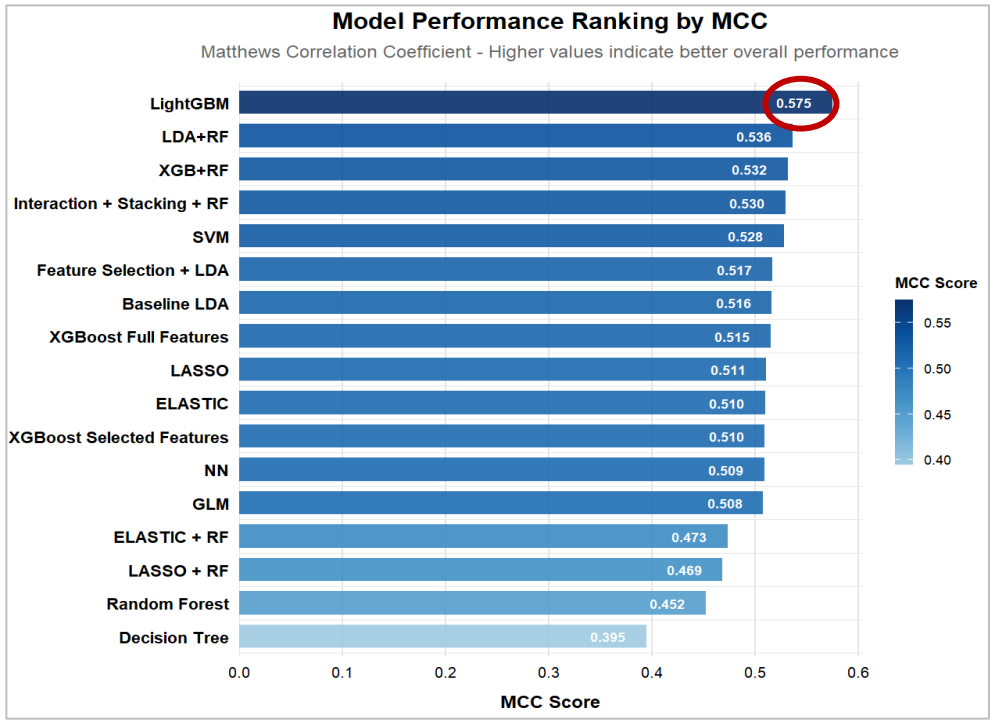
Recall

"How many of the actual subscribers did we successfully find?"

It's about capture rate.



Precision vs Recall scatter graph



MCC bar chart

A. High-Recall Models (Customer Acquisition Focus)—— LASSO — Maximizes customer coverage at the cost of more ineffective calls.

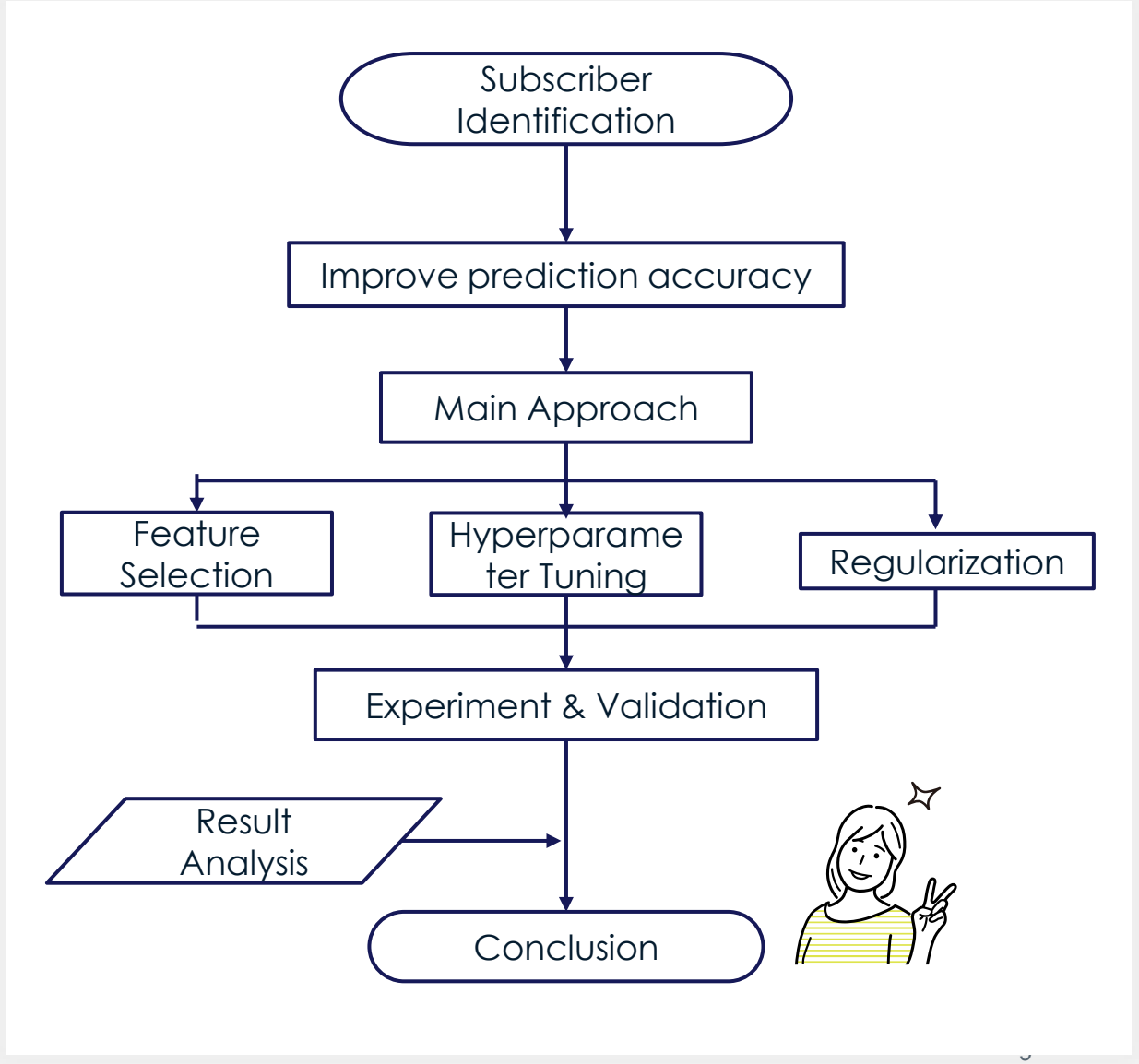
B. High-Precision Models (Cost-Control Focus)—— ELASTIC+RF — Ensures reliable predictions but may miss some potential subscribers.

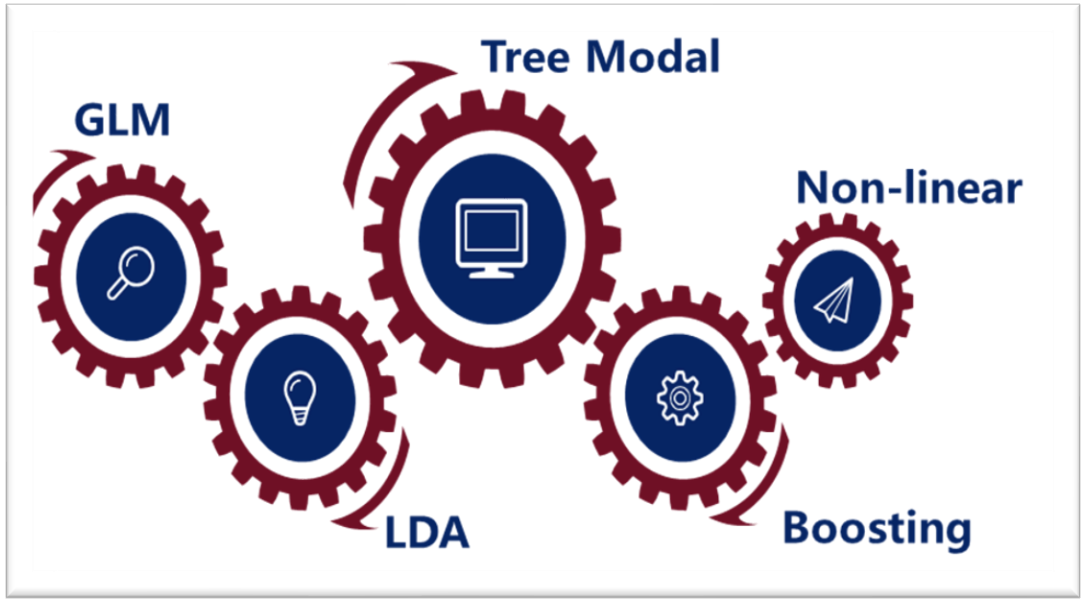
C. Balanced Models (Optimal Overall Performance)——LightGBM — Achieves optimal equilibrium between coverage and accuracy for real-world deployment.

Scenario	Objective	Top Model	Key Metric
Market Launch	Max Customer Reach	LASSO	Recall > 0.8
Standard Operations	Balanced Performance	LightGBM	MCC > 0.57
Resource-Rich Period	Maximize Performance	LightGBM / XGB+RF	MCC > 0.57
Cost-Sensitive Period	High Conversion Rate	ELASTIC + RF	Precision > 0.66
Stable Expansion	Balance & Interpretability	LDA + RF	MCC > 0.53

Final Recommendation: LightGBM

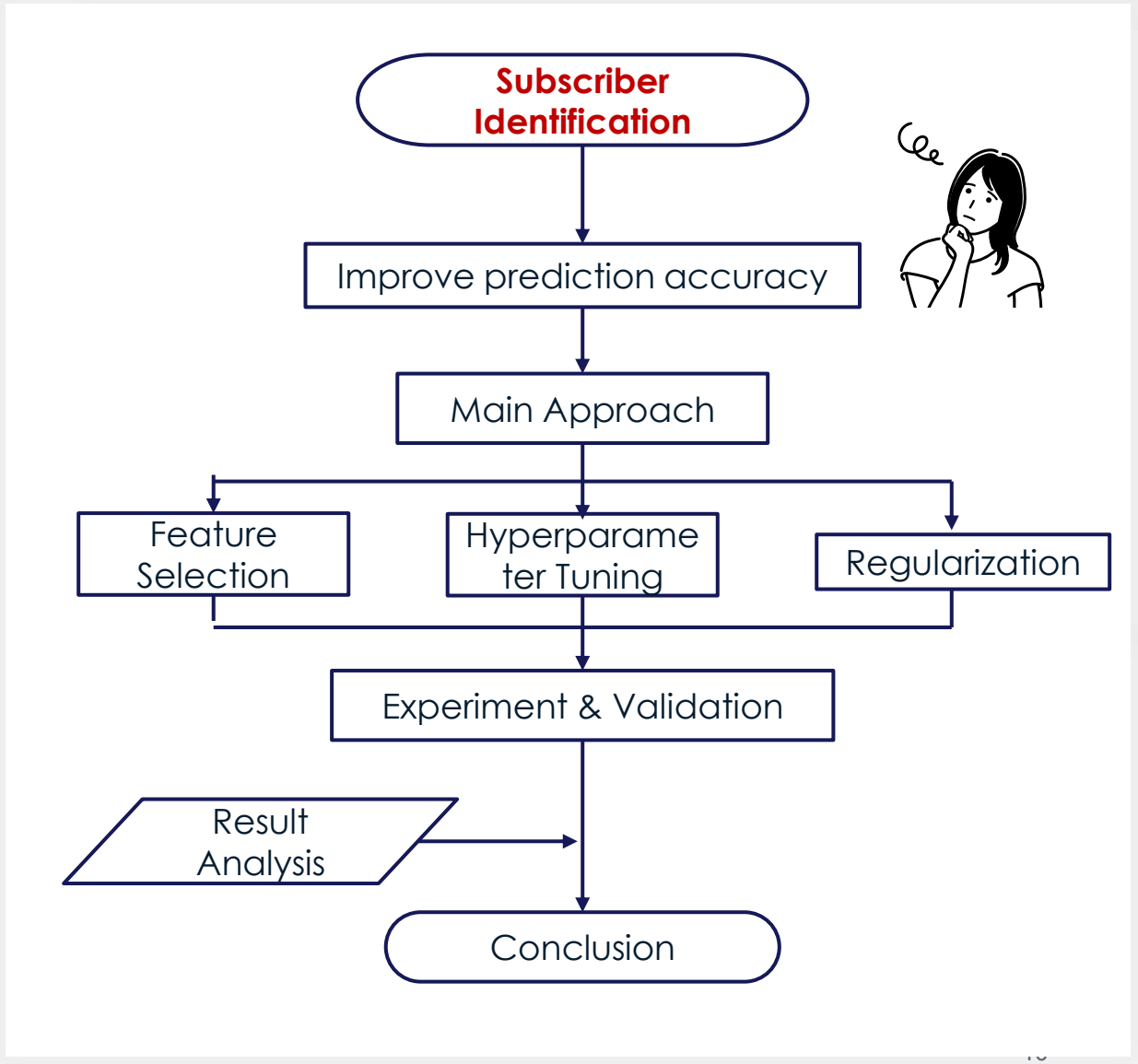
- Leading MCC and well-balanced performance.
- Delivers the most robust and reliable results for deployment.





Interpretability

High Performance



Summary

- Prioritize customers with long call durations.
- Target mobile users.
- Use LightGBM to capture the most value.

Thank you for listening!

Presenter:[ZHANG JINGYI]

Programe Title Here