CUSTOMER CHURN PREDICTION PROJECT
PHASE 3 – MACHINE LEARNING

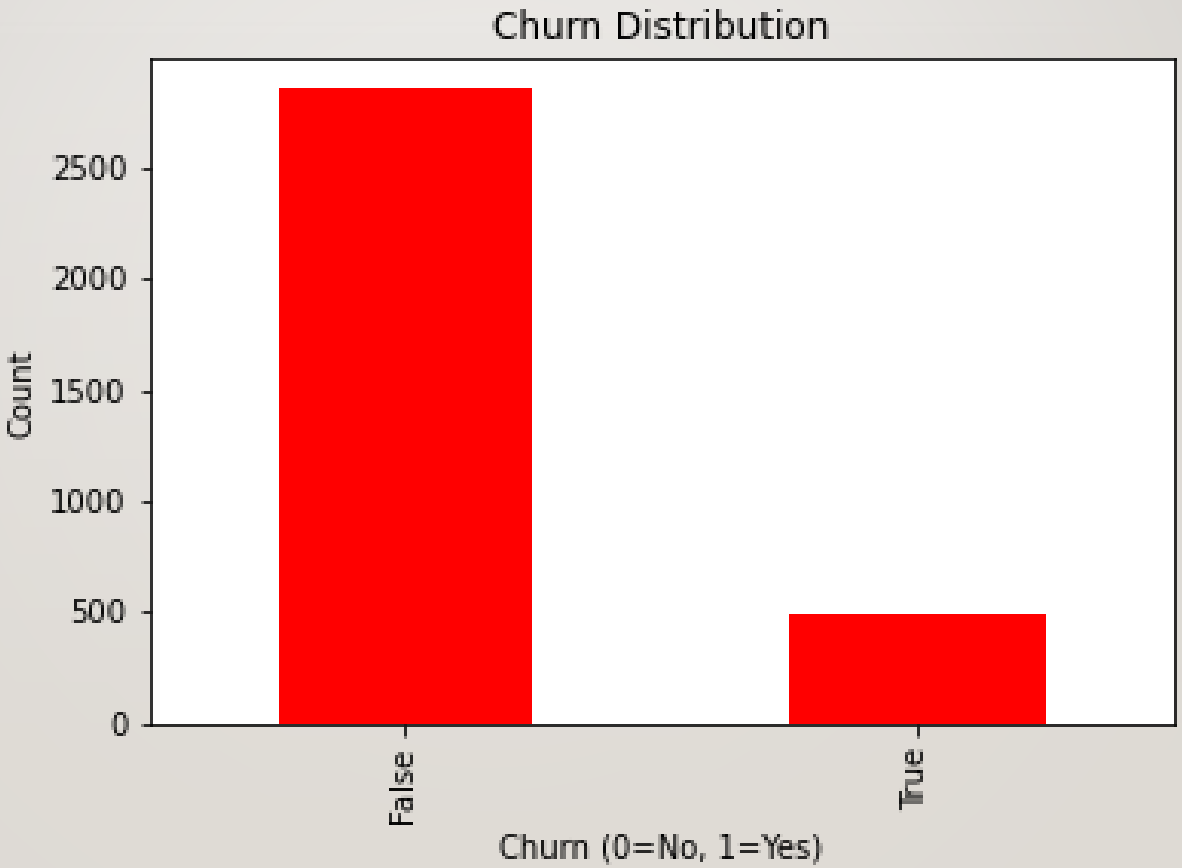
OBJECTIVE: IDENTIFY CUSTOMERS AT RISK OF LEAVING BEFORE THEY CHURN

TEAM: CUSTOMER SUCCESS & RETENTION ANALYTICS

BUSINESS PROBLEM & GOAL

- **Business Problem:**
- ✓ Losing customers reduces revenue and increases acquisition costs
- ✓ Retaining customers is more cost-effective than acquiring new ones
- **Goal:**
- ✓ Use data to proactively identify at-risk customers
- ✓ Enable retention teams to take targeted actions

GRAPHICAL PRESENTATION OF CUSTOMER CHURNING VS NOT CHURNING



MODELS TO BE TESTED

Model	Type	Notes
Logistic Regression	Linear	Baseline, interpretable
Decision Tree	Rule-based	Easy to explain
Random Forest	Ensemble	High precision
Gradient Boosting	Ensemble	Best balance of recall & precision

KEY METRICS TO BE USED TO MEASURE SUCCESS

- **Accuracy:** Overall correctness of the model
- **Precision:** When we flag a customer as at risk, how often are we correct?
- **Recall:** Out of all customers who actually churn, how many did we correctly identify? *(Most important for business)*
- **F1-Score:** Balance between precision and recall
- **AUC:** Overall ability to separate churners from non-churners

ACCURACY_SCORE COMPARISON

Model	Accuracy
Logistic Regression	85.5%
Decision Tree	88.8%
Random Forest	93.2%
Gradient Boosting	93.2%

RECALL_SCORE COMPARISON

Model	Recall
Logistic Regression	17.8%
Decision Tree	59.5%
Random Forest	57.9%
Gradient Boosting	61.6%

Business Insight:

Catching more at-risk customers prevents revenue loss

Gradient Boosting identifies the largest share of churners

PRECISION (RELIABILITY OF ALERTS)

Model	Precision
Logistic Regression	50.6%
Decision Tree	61.8%
Random Forest	92.7%
Gradient Boosting	86.1%

Business Insight:

Random Forest is highly reliable when it flags churners

Fewer unnecessary retention actions

F1-SCORE (BALANCED PERFORMANCE)

Model	F1-Score
Logistic Regression	0.263
Decision Tree	0.606
Random Forest	0.712
Gradient Boosting	0.718

Business Insight:

Gradient Boosting provides the best overall balance between detecting churners and minimizing false alarms

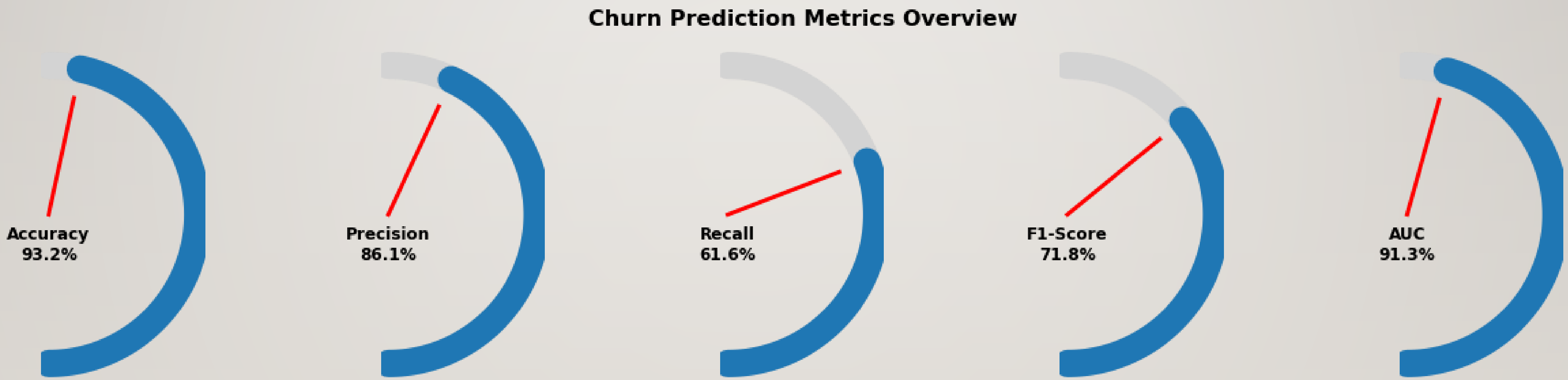
AUC (OVERALL MODEL STRENGTH)

Model	AUC
Logistic Regression	0.830
Decision Tree	0.766
Random Forest	0.906
Gradient Boosting	0.913

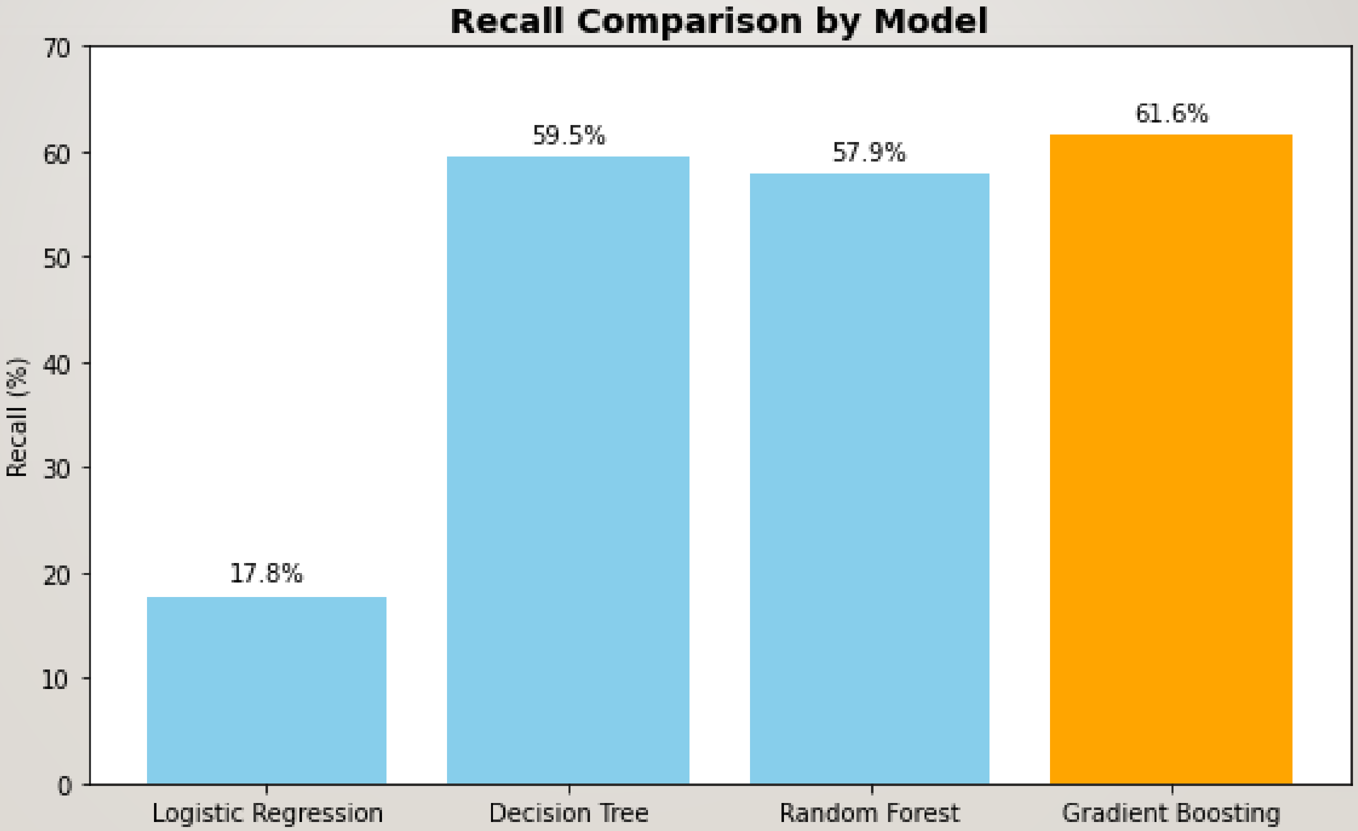
Business Insight:

Gradient Boosting is the strongest at distinguishing churners from non-churners

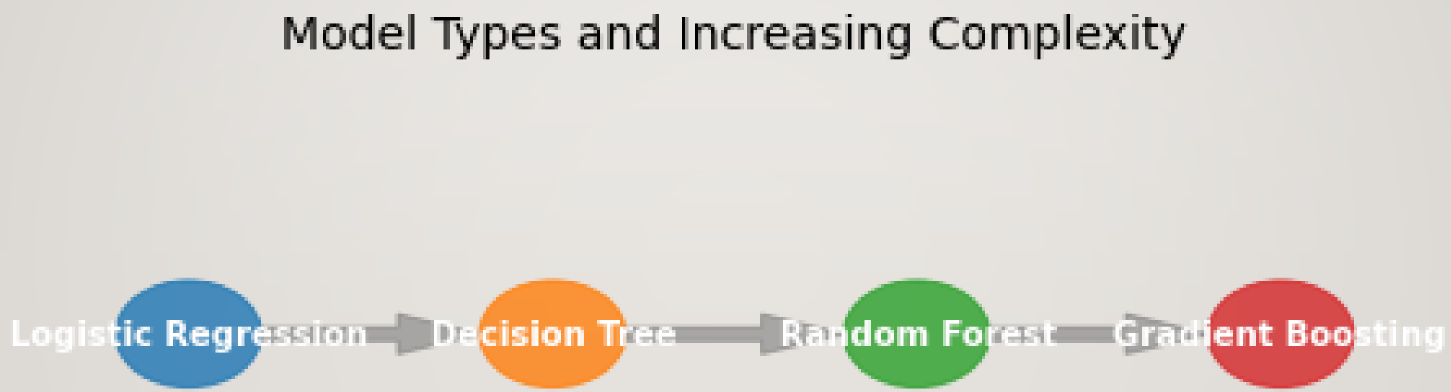
CHURN PREDICTION METRICS OVERVIEW



RECALL COMPARISON BY MODEL



MODEL TYPE AND INCREASING COMPLEXITY



LIMITATIONS

Class Imbalance: Some churners may still be missed

Threshold Dependence: Results vary depending on the probability cut-off

Feature Limitations: Missing external factors like competitor offers or customer sentiment

Complexity: Ensemble models are less interpretable than simple Decision Trees

Overfitting Risk: High-performance models may need careful monitoring for generalization

RECOMMENDATIONS

1. **Deploy Gradient Boosting** for proactive churn identification
2. **Rank high-risk customers** by predicted probability for targeted retention
3. **Integrate predictions into CRM** for real-time alerts
4. **Adjust thresholds** as business priorities shift
5. **Explore future enhancements**: additional features, resampling, or alternative ensembles

BUSINESS IMPACT

1. Identify ~62% of churners before they leave
2. Prioritize high-risk customers for retention campaigns
3. Reduce revenue loss and increase customer lifetime value
4. Enable data-driven, proactive decision-making