WHICH MODEL SHOULD I USE?

Model Approach Recommendation: Standalone vs Pipeline

Executive Summary

Recommendation: Use Standalone Prediction Model as Primary Approach

The standalone prediction model provides better coverage, more actionable insights, and simpler deployment. Use the pipeline approach (crosssale + prediction) only for the highest-confidence tier of leads.

Data Analysis

Standalone Model Results

- **Total Records**: 221,030
- **Coverage**: 100% of population
- **Distribution**:
  - `below_0_5` : 28.7% (63,448 records) - Low confidence
  - `above_or_equal_0_5` : 71.3% (157,582 records) - Moderate to high confidence
  - `above_or_equal_0_8` : 30.0% (66,284 records) - High confidence
  - `above_or_equal_0_9` : 11.2% (24,797 records) - Very high confidence
  - `above_or_equal_0_95` : 0.87% (1,926 records) - Extremely high confidence

Pipeline Model Results (Crosssale + Prediction)

- **Total Records**: 16,515 (7.5% of original population)
- **Coverage**: Only crosssale "yes" records
- **Distribution**:
  - `cross<0.5_pred<0.5` : 64.6% (10,673) - Both models say no/low
  - `cross>=0.5_pred>=0.5` : 9.7% (1,604) - Both models agree yes
  - `cross>=0.5_pred<0.5` : 54.7% (9,036) - **Crosssale says yes, prediction says no** ⚠
  - `cross>=0.8_pred<0.5` : 19.5% (3,223) - **Crosssale says yes, prediction says no** ⚠
  - `only_cross>=0.9` : 28.2% (4,667) - High crosssale, no prediction result
  - `cross<0.5_pred>=0.5` : 12.5% (2,073) - Crosssale says no, prediction says yes
  - `cross>=0.8_pred>=0.5` : 4.1% (680) - High crosssale with prediction
  - `cross>=0.8_pred>=0.8` : 0.036% (6) - **Both models high confidence**

Key Insights

1. Coverage Gap

- Pipeline approach filters out **92.5% of the population** (204,515 records)
- You're missing potential opportunities from clients who might buy a second policy but weren't flagged by crosssale model

2. Model Disagreement

- **54.7%** of pipeline records show crosssale model says "yes" but prediction model says "no/low confidence"
- This suggests the models are looking at different signals
- Only **0.036%** (6 records) have both models in high agreement

3. Actionable Volume

- Standalone: **66,284 records** with pred >= 0.8 (high confidence)
- Pipeline: **680 records** with cross>=0.8 AND pred>=0.5 (high confidence)
- **Standalone provides 97x more high-confidence leads**

4. False Negatives Risk

- **12.5%** of pipeline records show cross<0.5 BUT pred>=0.5
- These are clients the crosssale model missed but prediction model identified
- Using only pipeline would miss these opportunities

Recommendation: Hybrid Approach

Primary Strategy: Standalone Model

Use for: All clients (221,030 records)

Thresholds:

- **Tier 1 (Highest Priority)**: pred >= 0.9 → 24,797 records (11.2%)
- **Tier 2 (High Priority)**: pred >= 0.8 → 66,284 records (30.0%)
- **Tier 3 (Medium Priority)**: pred >= 0.5 → 157,582 records (71.3%)

Advantages:

- ✅ Full population coverage
- ✅ Clear probability-based prioritization
- ✅ Simpler deployment and maintenance
- ✅ More actionable volume at each tier
- ✅ No dependency on crosssale model accuracy

Secondary Strategy: Pipeline Model

Use for: Validation of top-tier leads only

When to use:

- For Tier 1 clients (pred >= 0.9), check if they also have cross>=0.8
- This creates a "platinum tier" of highest-confidence leads
- Expected volume: ~0.2% of total population (very small, highest quality)

Advantages:

- ✅ Additional validation layer
- ✅ Highest confidence when both models agree
- ✅ Useful for A/B testing or special campaigns

Implementation Strategy

Phase 1: Deploy Standalone Model (Immediate)

1. Use standalone model for all 221,030 clients
2. Segment by probability thresholds:
    - Tier 1: pred >= 0.9 (24,797 clients)
    - Tier 2: pred >= 0.8 (66,284 clients)
    - Tier 3: pred >= 0.5 (157,582 clients)
3. Prioritize outreach based on tiers

Phase 2: Add Pipeline Validation (Optional)

1. For Tier 1 clients, check crosssale model scores
2. Create "Platinum Tier": pred >= 0.9 AND cross >= 0.8
3. Use for special high-touch campaigns

Phase 3: Monitor and Optimize

1. Track conversion rates by tier
2. Adjust thresholds based on business results
3. Consider retraining if model disagreement is too high

Business Impact Comparison

Metric	Standalone Model	Pipeline Model
Total Leads	221,030	16,515
High Confidence (>=0.8)	66,284 (30%)	680 (4.1% of filtered)
Very High Confidence (>=0.9)	24,797 (11.2%)	~200 (estimated)
Coverage	100%	7.5%
False Negatives	Lower risk	Higher risk (misses 92.5%)

Metric	Standalone Model	Pipeline Model
Deployment Complexity	Simple	More complex (2 models)

Conclusion

Use the Standalone Prediction Model as your primary approach because:

1. **Better Coverage**: 13x more leads (221K vs 16.5K)
2. **More Actionable**: 97x more high-confidence leads (66K vs 680)
3. **Lower Risk**: Captures opportunities the crosssale model might miss
4. **Simpler**: Single model to deploy and maintain
5. **Flexible**: Probability thresholds allow dynamic prioritization

The pipeline approach is valuable only as a secondary validation layer for the highest-tier leads, not as the primary strategy.