# Week 12 - Save and package your model for deployment

**Maria Alice Fagundes Vieira**

**Aritra Ray**

**Environment dependencies**

Operating System: Windows 10

OS Version: 10.0.19045

Python Version: 3.11.5

Pandas version: 2.2.2

NumPy version: 1.24.3

Seaborn version: 0.13.1

Matplotlib version: 3.8.0

Scikit-learn version: 1.3.0

Joblib version: 1.3.2

**Batch vs. Real-Time Inference**

The model will be deployed in batch mode due to its nature as a monthly recommendation system. Since the goal is to predict which products clients will purchase in the following month, real-time updates are unnecessary. However, to address new clients, batch predictions could run daily to provide immediate recommendations for those who recently opened accounts. For existing clients, recommendations can be updated biweekly or monthly, allowing time for client outreach and engagement. This frequency balances computational efficiency with responsiveness, ensuring the bank is proactive in offering relevant products.

**Performance Metrics and Monitoring Plan**

Key performance metrics to monitor include:

1. **Model Metrics**:

- **ROC AUC**: Measures the model's ability to distinguish between classes (important for assessing overall performance).
- **F1 Score**: Balances precision and recall, highlighting the trade-off between recommending the right product and avoiding incorrect recommendations.

2. **Business Metrics**:

- **Conversion Rate**: Fraction of recommended products that match actual purchases.
- **Error Rate**: Proportion of products purchased but not recommended, indicating missed opportunities.
- **Negative Hits**: Instances where irrelevant recommendations frustrate customers, potentially harming relationships.

3. **Customer Feedback**:

- Feedback-based insights or behavioral changes (e.g., reduced engagement) can signal dissatisfaction with recommendations.

**Thresholds for Monitoring Metrics**

Based on current metrics (Avg ROC AUC: 0.884, Avg F1 Score: 0.212), good thresholds for monitoring are defined as the below and provide a structured way to identify when to take corrective action or pull the model from production. The thresholds were chosen based on minor fluctuations for green flags, moderate deviations for yellow flags that reflect tolerable underperformance, signaling a need for closer tracking without immediate disruption and critical underperformance for red flags that compromises recommendation quality, likely leading to missed opportunities or customer dissatisfaction. At this point, the model may no longer deliver business value and requires retraining or removal.

**Green Flags**:

- ROC AUC: ≥ 0.85
- F1 Score: ≥ 0.20
- Conversion Rate: ≥ 15%
- Error Rate: ≤ 5%
- Negative Hits: ≤ 2%

**Yellow Flags**:

  - ROC AUC: 0.75 - 0.85
  - F1 Score: 0.15 - 0.20
  - Conversion Rate: 10% - 15%
  - Error Rate: 5% - 10%
  - Negative Hits: 2% - 5%

- **Red Flags**:
  - ROC AUC: < 0.75
  - F1 Score: < 0.15
  - Conversion Rate: < 10%
  - Error Rate: > 10%
  - Negative Hits: > 5%

**Risk Mitigation Strategies**

- **Green Flags**: Continue monitoring regularly. Ensure infrastructure and processes are running smoothly.
- **Yellow Flags**: Investigate potential causes such as data drift, feature importance shifts, or seasonal patterns. Perform a more frequent review of input data distributions and feature behavior.
- **Red Flags**: Temporarily remove the model from production and retrain it with updated data. Evaluate training data, parameters, and potential data or concept drift. Engage in a deeper analysis of misclassifications and business outcomes.

**Retraining Frequency**

The model will be retrained monthly. This schedule aligns with the availability of new data from client transactions and behavior changes, ensuring the model remains adaptive to evolving customer needs. Monthly retraining also captures seasonal trends and mitigates the impact of data or concept drift.

**Data Drift, Concept Drift and Mitigation Strategies**

Both data drift and concept drift are potential risks. Data drift could represent changes in client demographics, such as shifts in age or income distributions, and concept drifts could be changes in product preferences due to external factors like economic fluctuations or new product launches.

Mitigation strategies would include monitoring data distributions regularly using the EDA techniques we used back on week 3 and check distributions on the data. Additionally, we could use track statistical properties of features over time. For concept drift, we could monitor the relationship between recommendations and actual purchases as well as flag substantial deviations from historical trends.