IMT Atlantique
Bretagne-Pays de la Loire
École Mines-Télécom

Evaluation: From metrics to business impact

Data Science Toolkit & Applications
Lina Fahed & Laurent Brisson

Evaluation in CRISP-DM: The Final Test 🎯

Evaluation: Measuring the Power of Your Model

From Metrics to Business Impact in CRISP-DM

Core Objective:

- Ensure the **model meets business goals** and performs well on unseen data.

Example:

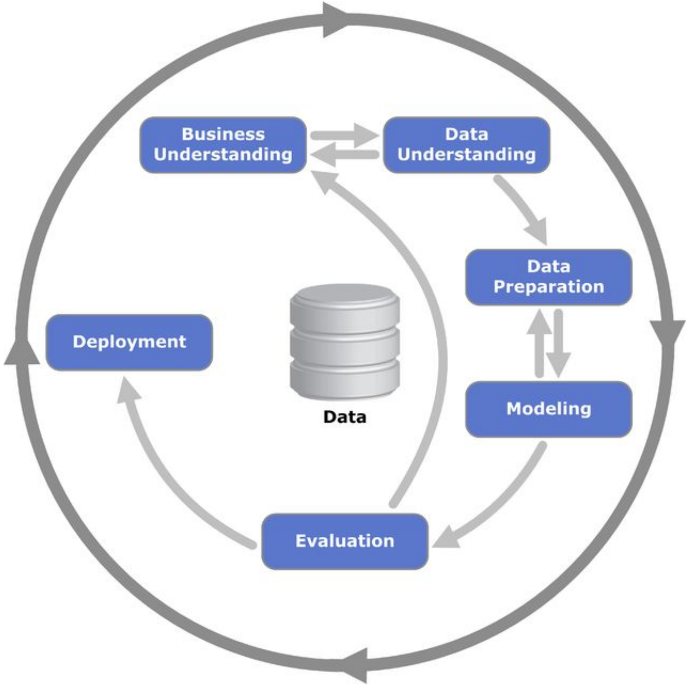
- A churn prediction model with 95% accuracy is useless if it doesn't reduce customer attrition.

"Not everything that can be counted counts, and not everything that counts can be counted." — Albert Einstein

Where Does Evaluation Fit?

CRISP-DM Steps:

1. Business Understanding ✅
2. Data Understanding ✅
3. Data Preparation ✅
4. Modeling ✅
5. **Evaluation** (Current focus) 🔍
6. Deployment



Evaluation Goal:

- Assess **model performance**
- **align with business objectives**.

What Does Evaluation Involve?

- **Review Results**: Check model performance on test data. 📊
- **Assess Business Impact**: Does the model solve the problem? 💼
- **Validate with Stakeholders**: Get feedback from domain experts. 🗣️
- **Iterate or Proceed**: Decide if the model is ready for deployment or needs improvement. 🔄

Technical Evaluation – Metrics / Measuring Model Performance 📈

Problem Type	Metrics	When to Use
Classification	Accuracy, Precision, Recall, F1-Score, ROC-AUC, Confusion Matrix	Spam detection, churn prediction
Regression	MAE, MSE, RMSE, R^2	House price prediction, sales forecasting
Clustering	Silhouette Score, Davies-Bouldin Index, Elbow Method	Customer segmentation
Association	Support, Confidence, Lift	Market basket analysis

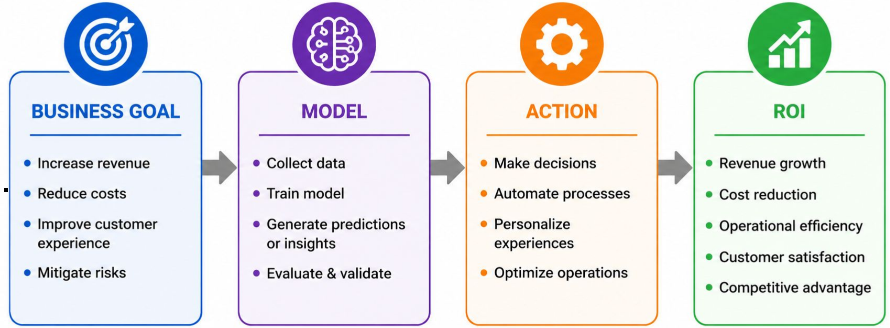
Business Evaluation – Aligning with Goals
Does the Model Solve the Business Problem? 💼

Key Questions:

- Does the model meet the success criteria? e.g., reduce churn by 10%
- Is the model actionable? Can business teams use the predictions?
- What is the ROI? Cost of implementation vs. benefits.

Example:

- **Business Goal**: Reduce customer churn by 10%.
- **Model Impact**: Predicts churn with 90% accuracy
- **Action**: Target high-risk customers with retention offers.



Remember:

- Business Goal: What business problem are we trying to solve?
- Model: AI/ML model that generates predictions or recommendations.
- Action: Decisions or processes driven by the model's output.
- ROI: Measurable business value (revenue, savings, productivity, customer satisfaction, etc.).

Confusion Matrix – Deep Dive
Understanding Classification Errors 🔍

Derived Metrics:

- **Accuracy**: (TP + TN) / Total
- **Precision**: TP / (TP + FP)
- **Recall**: TP / (TP + FN)
- **F1-Score**: 2 * (Precision * Recall) / (Precision + R

Confusion Matrix

	Actually Positive (1)	Actually Negative (0)
Predicted Positive (1)	True Positives (TPs)	False Positives (FPs)
Predicted Negative (0)	False Negatives (FNs)	True Negatives (TNs)

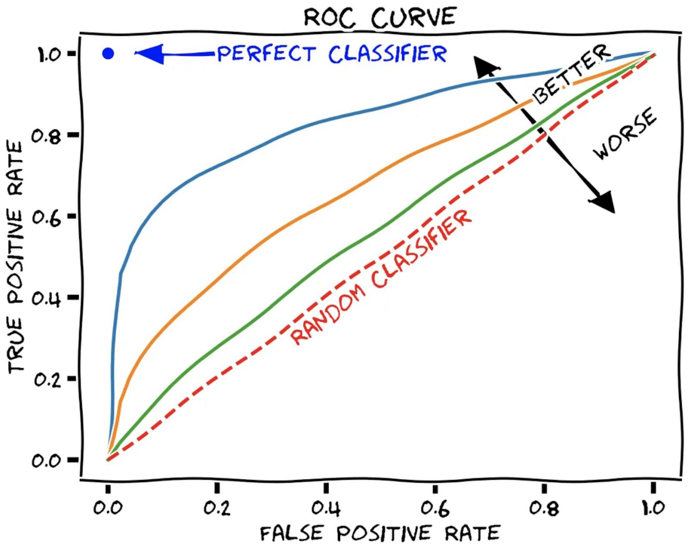
FN (false negatives) are often more costly than FP (false positives) in applications like fraud detection or medical diagnosis

Visualizing Model Trade-offs 📉

ROC Curve:
- Plots True Positive Rate (TPR) vs. False Positive Rate (FPR) at different thresholds.

AUC (Area Under Curve):
- AUC = 1.0: Perfect model.
- AUC = 0.5: Random guessing.
- AUC > 0.7: Generally acceptable.

Example:
- A model with AUC = 0.9 can distinguish classes well.



IMT Atlantique
Bretagne-Pays de la Loire
École Mines-Télécom

Choosing the Best Model 🏆

Model	Accuracy	Precision	Recall	F1-Score	Training Time	Interpretability
Logistic Regression	85%	82%	80%	81%	Fast	High
Random Forest	88%	85%	80%	82%	Medium	Medium
XGBoost	89%	86%	81%	83%	Slow	Low
Neural Network	90%	87%	82%	84%	Very Slow	Very Low

Recommendation:

- **Best Performance**: Neural Network (but slow and hard to interpret).
- **Best Trade-off**: XGBoost (high performance, reasonable speed).
- **Best for Interpretability**: Logistic Regression.

Stakeholder Validation Getting Buy-In from the Business 🤝

Why Validate with Stakeholders?

- Ensure the model aligns with business needs and is trusted by end-users.

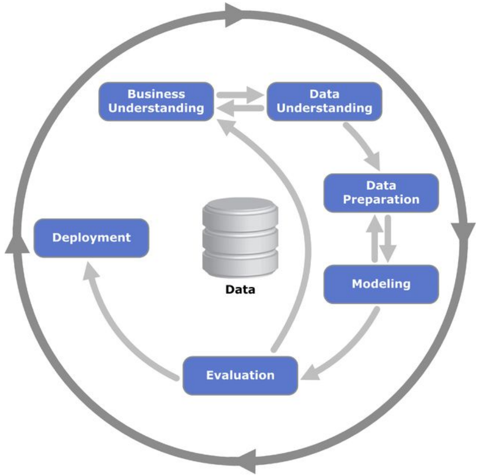
How to Present Results:

- Avoid Technical Jargon
  - Use business language
  - e.g., 'This model can reduce churn by 10%'
- Show Impact
  - Demonstrate ROI
  - e.g., cost savings, revenue increase
- Address Concerns
  - Be transparent about limitations.

Example Presentation:

"Our churn model identifies 80% of at-risk customers, allowing us to reduce churn by 10% with targeted retention offers."

Next Steps: Iterate or Deploy? 🔄



When to Iterate:

- Model performance is below business goals.
- Stakeholders reject the model (e.g., low trust, poor alignment with needs).
- New data or business requirements emerge.

When to Deploy:

- Model meets success criteria and is actionable.
- Stakeholders approve the model.
- The ROI justifies deployment.

Example:

- If the churn model's recall is 80% but the business needs 90%, iterate by collecting more data or trying new algorithms.

Essential Tools and Libraries 🛠️

Python Libraries:

- Scikit-learn: accuracy_score, precision_score, confusion_matrix
- Matplotlib/Seaborn: for ROC curves, confusion matrices
- Yellowbrick: for visual diagnostics

Other Tools:

- R (caret, MLmetrics)
- Excel (for business metrics)
- Tableau/Power BI (for visualizing impact)

What to Remember

- Evaluation is More Than Metrics – Align technical performance with business goals. 💼
- Stakeholders Matter – A model is only successful if the business uses and trusts it. 🤝
- Iterate or Deploy – Use evaluation results to decide the next steps. 🔄

IMT Atlantique
Bretagne-Pays de la Loire
École Mines-Télécom