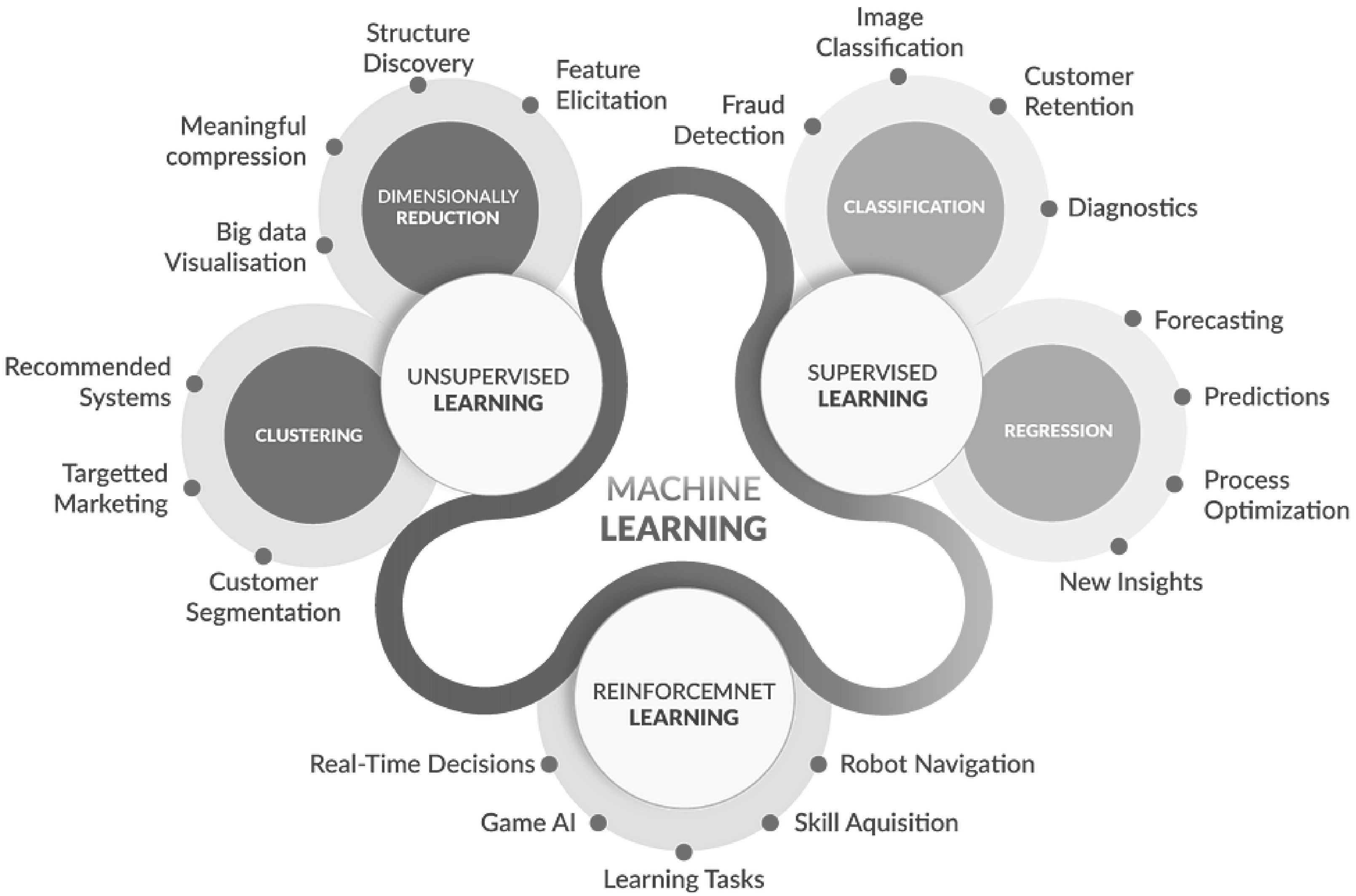
Smarter Marketing, Stronger Results:
Leveraging Machine Learning in Banking

Machine Learning Marvels: **Exploring Different Learning Styles**



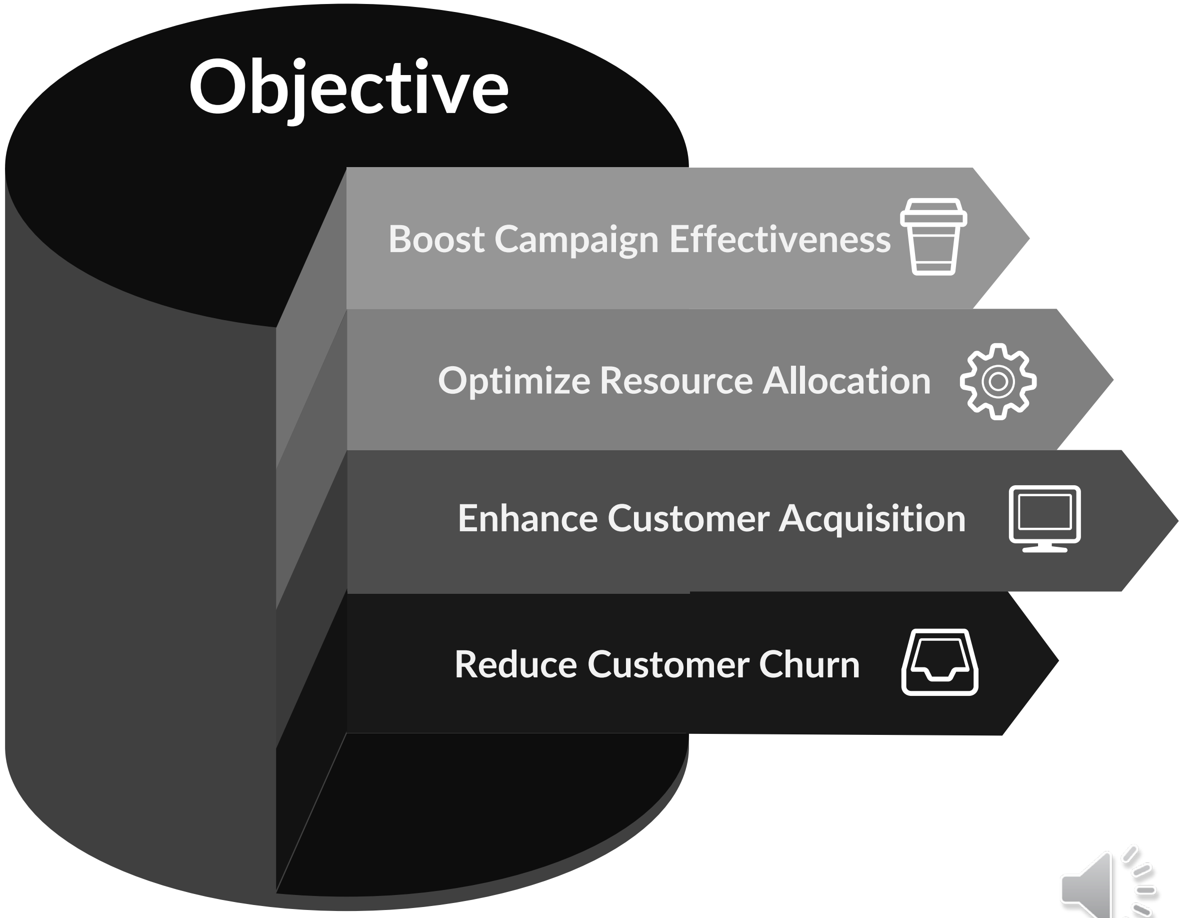
Machine Learning: The Invisible Force Shaping Our World

1. **Speech Recognition**
   Classifies spoken words into text (e.g., Siri and Alexa).
2. **Spam Filtering**
   Classifies emails as spam or not spam
3. **Personalized News Feeds**
   Recommends news articles based on past reading history (classification)
4. **Traffic Prediction**
   Identifies patterns in traffic flow data to predict congestion (clustering).
5. **Self-Driving Cars**
   Learns optimal driving behavior through trial and error (reinforcement)
6. **Fraud Detection**
   Identifies suspicious transactions by learning patterns of legitimate transactions
7. **Recommendation Systems**
   Recommends products based on user behavior and similar user preferences
8. **Medical Diagnosis**
   Assists doctors by identifying patterns in medical images that might indicate abnormalities
9. **Natural Language Processing**
   Tasks like machine translation
10. **Customer Segmentation**
    Groups customers with similar characteristics based on their purchase history, demographics, or website behavior

Predicting Customer Response: **Optimizing Bank Marketing Campaigns**

Scope

This project focuses on applying machine learning techniques to analyze a dataset containing customer data from a recent marketing campaign conducted by a bank. The goal is to develop a predictive model that can anticipate customer response (subscription) to the bank's new product offering.



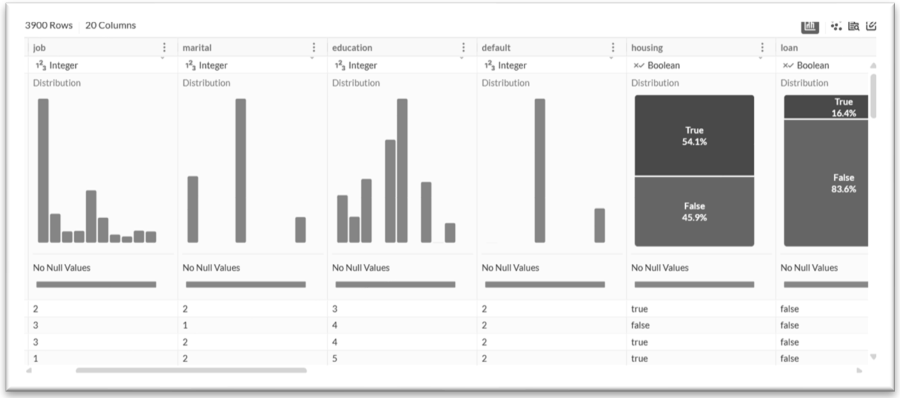
Behind the Scenes: **Decoding Customer Data**

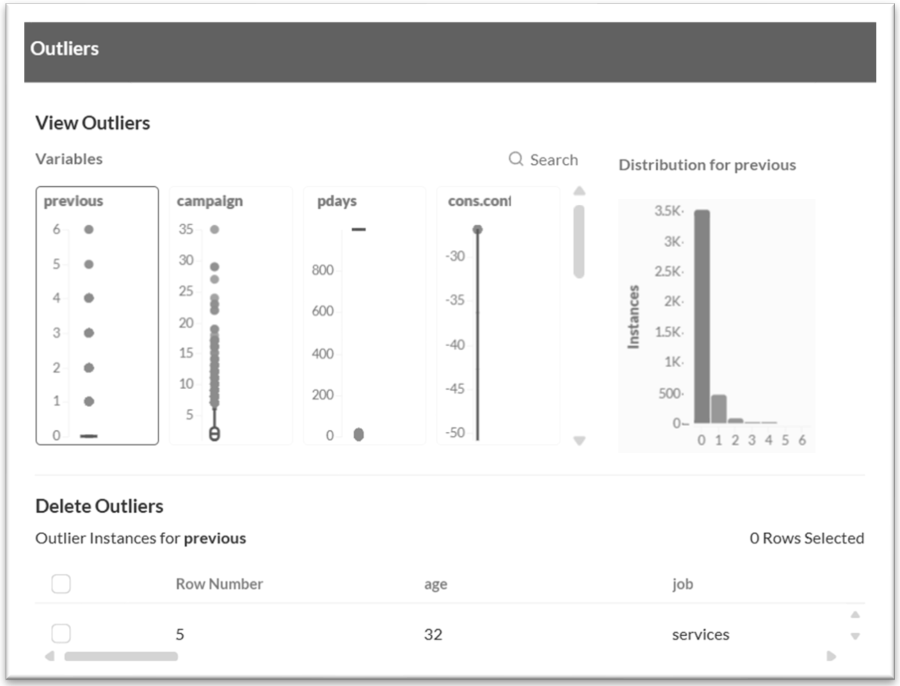
The analysis utilizes historical customer data from the bank's marketing campaigns. It leverages various key data elements

Category	Data Element	Category	Data Element
Customer Demographics	Age, Job	Previous Contact History	Days Since Previous Contact
Customer Demographics	Education, Marital Status	Previous Contact History	No of Previous Contacts
Financial Information	Previous Credit Default	Previous Contact History	Previous Contact Outcome
Financial Information	Having Housing Loan	Economic Indicators	Employment Variation Rate
Financial Information	Having a Personal Loan	Economic Indicators	Consumer Price Index
Marketing Campaign Details	Contact Method	Economic Indicators	Consumer Confidence Index
Marketing Campaign Details	Month of Contact, Day of Week	Economic Indicators	3-month Euribor Rate
Marketing Campaign Details	Campaign Type	Economic Indicators	Number of People Employed

Target Variable (y): Indicates if the customer subscribed to the offered product ("yes") or not ("no") after the marketing campaign. This is the variable the machine learning model will try to predict.

Transforming Raw Data: The Preprocessing Journey





Missing Values

Removing records with a significant number of missing values (115 out of 4119)

Outlier Detection

Outliers were identified, but removing them can introduce bias. To avoid this, the data were kept.

Formatting Consistency

It was ensured that all variables had consistent formatting

Encoding Categorical Variables

Categorical data like "marital status" was converted to numerical values (e.g., "single" = 1) for model compatibility

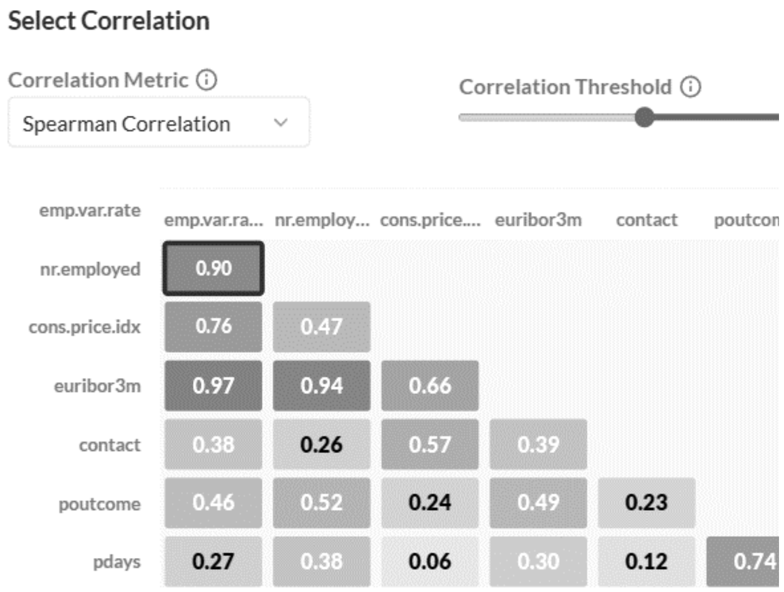
Unveiling Customer Behavior: **Data Exploration**

Exploring the bank marketing data is key to understanding customer behavior. Spearman's Rank Correlation, a method robust to outliers, is used to identify valuable relationships.

Insights

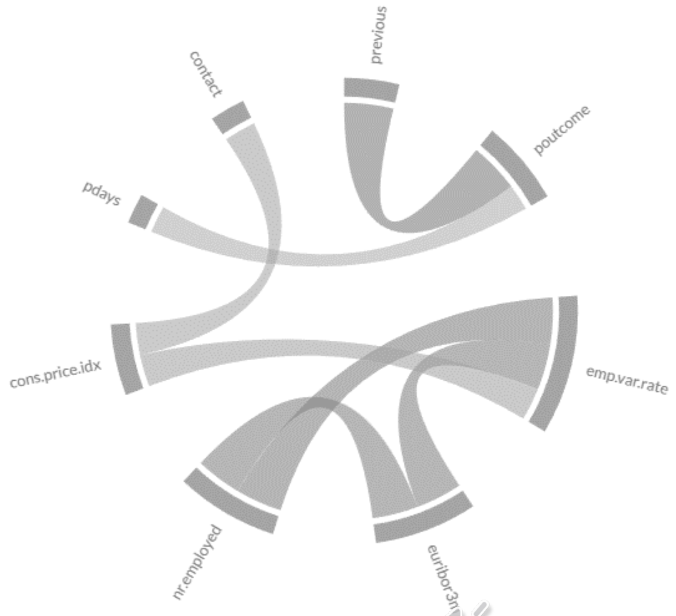
- Economic Indicators: Strong positive correlation (0.66-0.9) between employment, consumer price index, and interest rates suggests a healthy economy benefits marketing new products.
- Previous Contact and Campaign Success: Positive correlation (0.74) between days since last contact and campaign outcomes indicates customers not contacted recently might be receptive to new offerings.
- Leveraging insights help building a model and refining it with techniques like feature selection.

Correlation Matrix



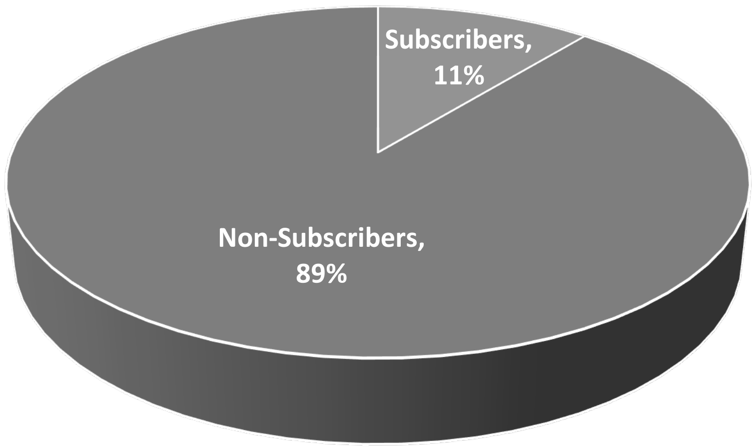
	emp.var.ra...	nr.employ...	cons.price....	euribor3m	contact	poutcon
nr.employed	0.90					
cons.price.idx	0.76	0.47				
euribor3m	0.97	0.94	0.66			
contact	0.38	0.26	0.57	0.39		
poutcome	0.46	0.52	0.24	0.49	0.23	
pdays	0.27	0.38	0.06	0.30	0.12	0.74

Chord Diagram



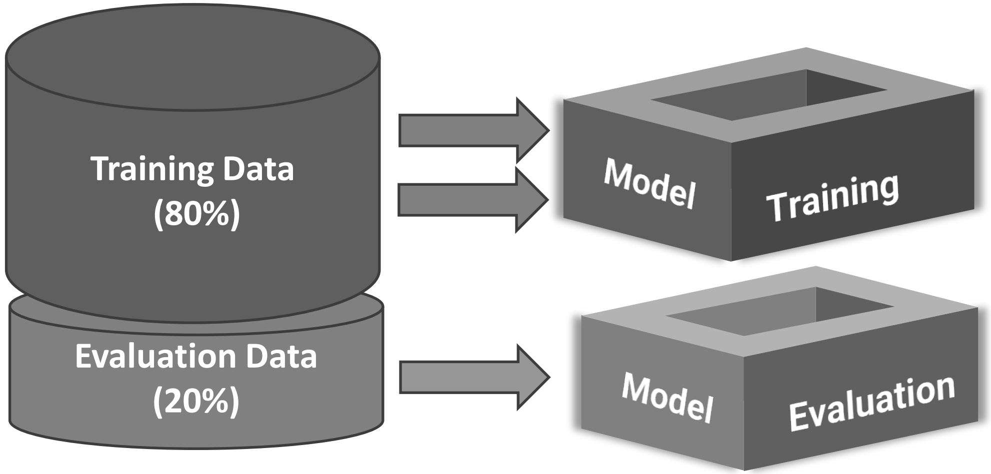
Training, Evaluating and Selecting **ML Model**

Why Classification

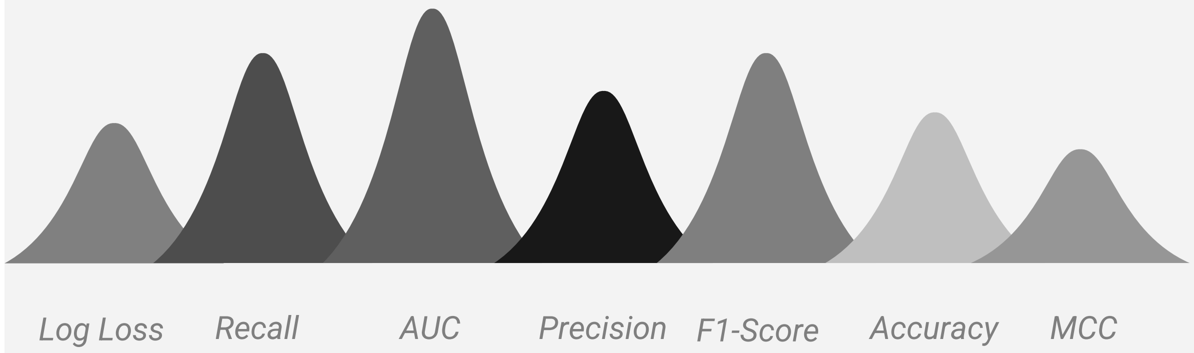


The aim is to categorizing customers as potential subscribers or not. This makes classification models a better choice.

Model Training and Evaluation

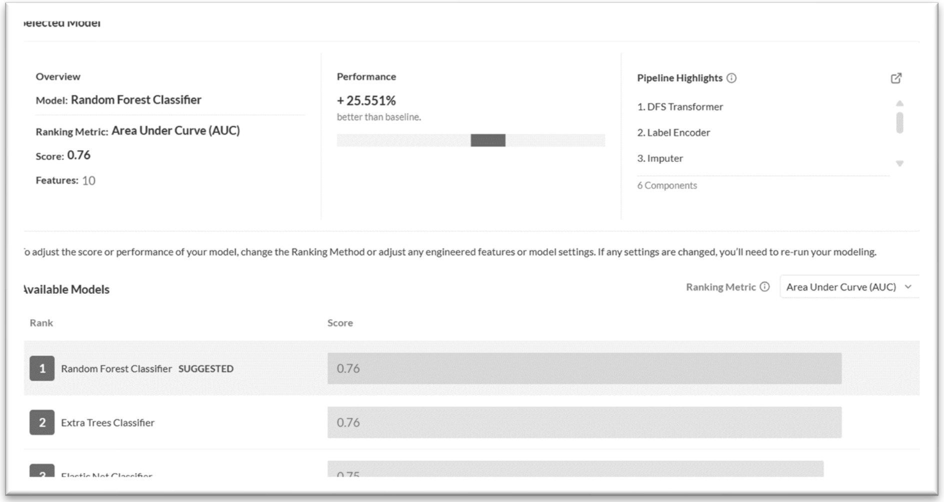


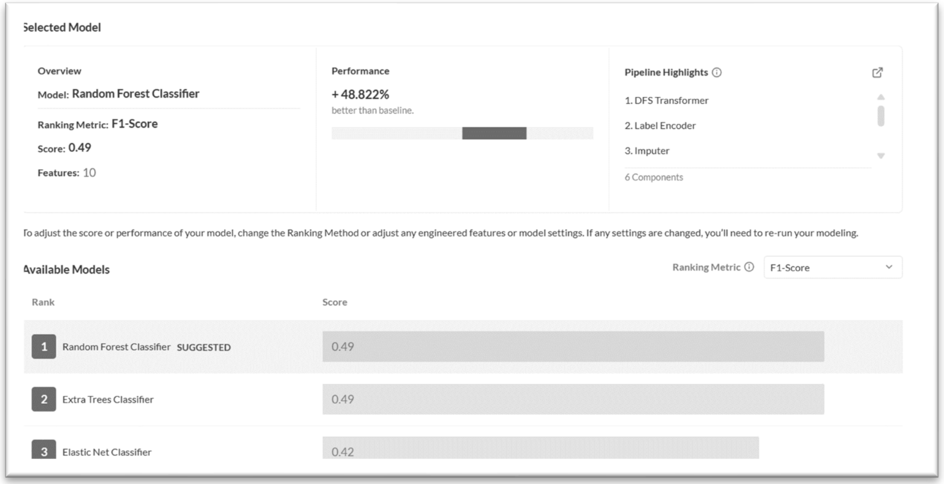
the purpose of the split is to prevents Overfitting



- Selecting the right evaluation metric is crucial for model selection and performance
- The chosen metric depends on data balance and model purpose
- Bank marketing data is likely imbalanced (more non-subscribers)
- **AUC** and **F1-Score** are generally recommended over accuracy

Unveiling the Best Model: **Evaluating Performance**





Evaluating Model Performance

AUC: Random Forest model achieved a score of 0.76, indicating good overall performance. The significant improvement over the baseline (25.55%) emphasizes its effectiveness.

F1-Score: Random Forest model's F1-Score of 0.49 suggests a decent balance between these factors. However, the lower improvement compared to the baseline (48.82%) might indicate room for improvement.

Interpreting the Results

AUC: As a robust metric for imbalanced data, the high AUC and significant improvement over the baseline suggest the Random Forest model excels at differentiating between classes.

F1-Score: The lower improvement compared to the baseline could indicate the model struggles with correctly identifying subscribers.

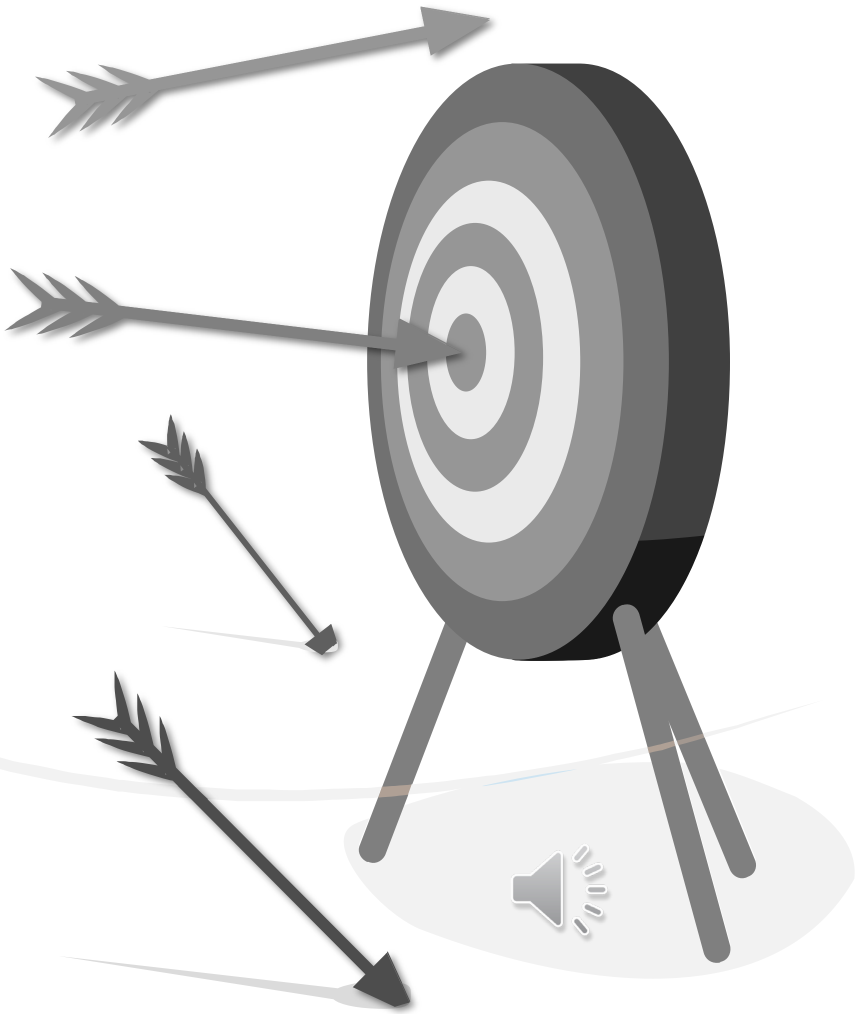
Based on the high AUC and significant improvement over the baseline in both metrics, the **Random Forest model** emerges as the strongest contender for overall performance. To evaluate generalizability, the model will be tested on a fresh dataset.

Data-Driven Business Decision Making

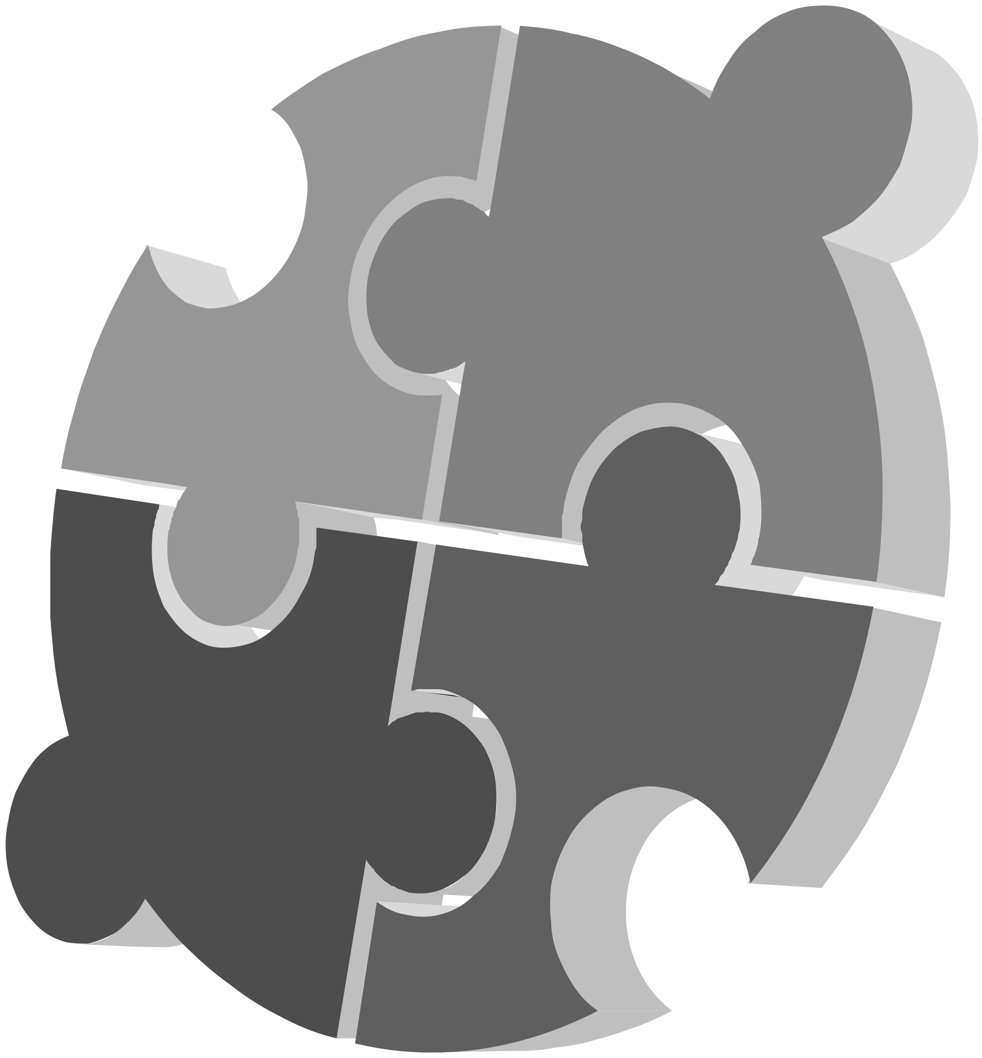
High-Propensity Subscribers: Our model identifies customers most likely to subscribe (high True prediction probability). These are ideal targets for marketing campaigns, allowing the bank to focus efforts and optimize budgets on those most likely to respond positively.

Model Performance: Our model achieves good overall **accuracy (80%)** in differentiating subscribers. High **precision (85%)** and **recall (90%)** indicate it effectively identifies potential subscribers and avoids wasted efforts. This ensures bank doesn't miss out on valuable customers who might respond to its campaigns.

Beyond High-Value Targets: Customers with lower prediction probability still hold potential. The bank can develop targeted marketing strategies to engage them.



Responsible AI in Banking: **Ethical & Legal Considerations**



Data Privacy

Complying with regulations like when collecting and using customer data. Robust security measures are essential.

Transparency & Explainability

Customers deserve to understand how their data is used. We have to explore techniques for model explainability to build trust and address concerns about "black box" algorithms.

Bias & Fairness

Mitigating potential biases in the training data to ensure model predictions are fair across different customer groups.

Consumer Protection

Marketing campaigns based on predictions should be transparent and avoid deceptive practices.

THANKS FOR WATCHING