# Predicting customer response

Report

Bill Meehan

# Overview

- **Problem Statement & Business Objectives**
  - Business goal: **Increase term deposit subscriptions** through targeted marketing
  - Challenge: **Identify customers most likely to subscribe** using available data
  - Objective: **Optimize marketing efforts** to improve conversion rates and ROI
- **Data Science Formulation**
  - Predict customer subscription likelihood using **machine learning models (customer propensity scoring)**
  - Transform business goal into a **binary classification problem**
  - Use customer demographics, campaign, and economic data as predictive features
- **Success Metrics**
  - **Accuracy**: Overall correctness of predictions
  - **Precision & Recall**: Balance between targeting true positives and minimizing false contacts
  - **AUC/ROC**: Model's ability to distinguish subscribers from non-subscribers
  - **Business ROI**: Increased conversion rate and reduced marketing costs

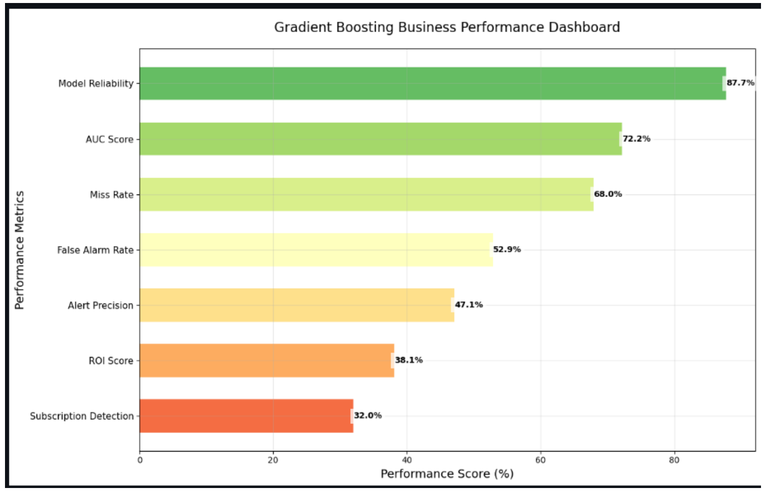
**Success Metrics**

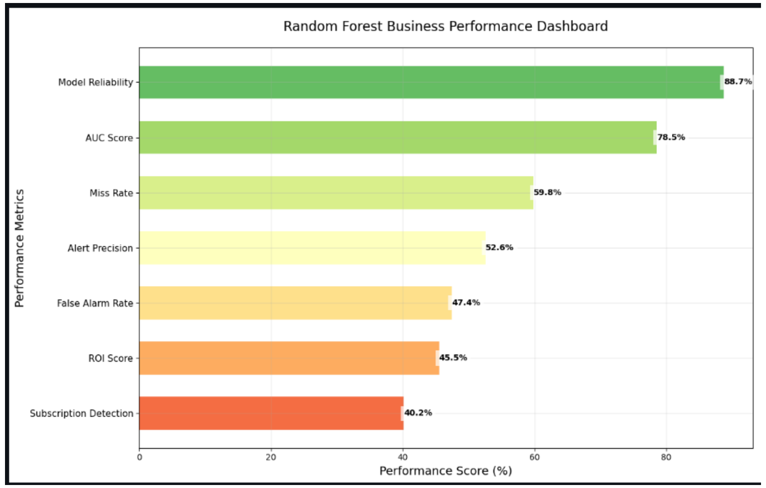
- Model Performance
  - Achieve AUC-ROC score > 0.85
  - Maintain precision > 60%
  - Achieve recall > 60%
  - F1 score > 0.60
- Business Impact
  - Increase conversion rate by 20%
  - Reduce marketing contacts by 30%
  - Lower customer acquisition cost by 25%
  - Improve campaign ROI by 40%
- Operational Metrics
  - Model inference time < 100ms per customer
  - Daily prediction updates
  - < 1% missing values in key features
  - Monthly drift monitoring reports

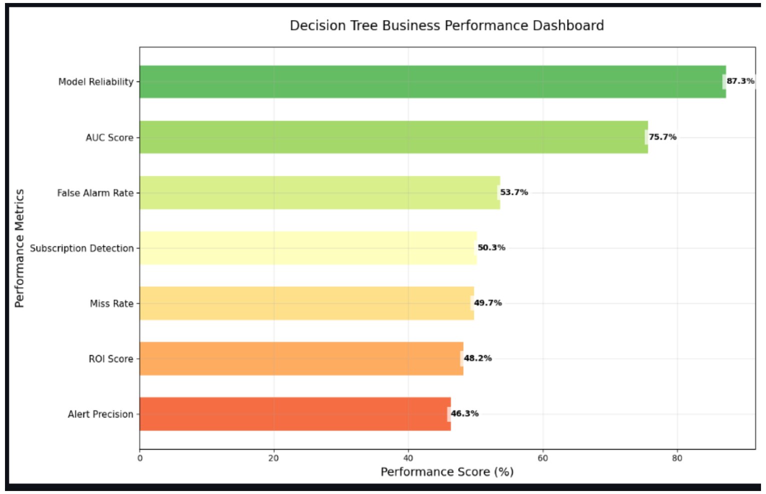
# Approach

- **Defined business objectives** to guide analysis and outcomes
- **Collected and prepared data** for robust, reliable insights
- **Explored data patterns** to identify key trends and drivers
- **Applied statistical modelling** to quantify relationships and predict outcomes
- **Interpreted results** to generate actionable recommendations

- **Statistical Modelling Concepts Used**
- **Regression analysis:** Measures impact of variables on outcomes
- **Classification models:** Categorises data for decision-making
- **Validation techniques:** Ensures model accuracy and reliability
- **Feature selection:** Identifies most important predictors
- **Visualisation:** Communicates complex results clearly




**Metric Descriptions:**

- Model Reliability: Overall accuracy in predictions
- Subscription Detection: Ability to identify customers likely to subscribe
- Alert Precision: Accuracy of subscription predictions
- False Alarm Rate: Percentage of incorrect subscription predictions
- Miss Rate: Percentage of missed subscription opportunities
- ROI Score: Balance of precision and recall
- AUC Score: Overall model discrimination

# Findings

**Data Quality & Preparation**

- Initial data showed significant class imbalance (88% "no", 12% "yes" for term deposit subscription).
- Detected and removed leaky features (e.g., 'duration') to prevent data leakage and overfitting.
- Applied SMOTE to balance classes, improving model fairness and learning.

**Feature Engineering**

- Created new features combining economic indicators, campaign effectiveness, and customer segmentation.
- One-hot encoding and imputation ensured robust handling of categorical and missing data.

**Model Performance**

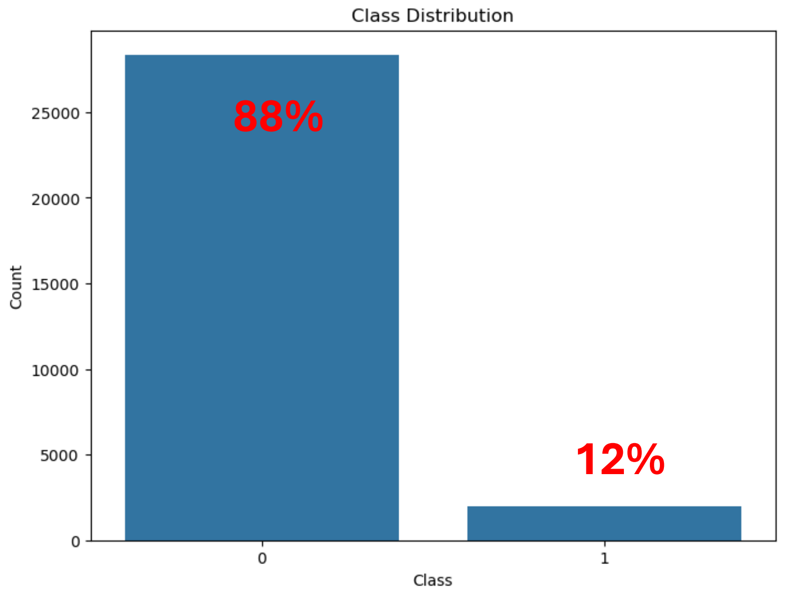
- All models (Decision Tree, Random Forest, GBM) achieved strong accuracy (DT: 87%, RF: 89%, GBM: 88%).
- Random Forest and GBM outperformed Decision Tree in AUC and F1-score, indicating better predictive power.
- Hyperparameter tuning further improved model stability and performance.

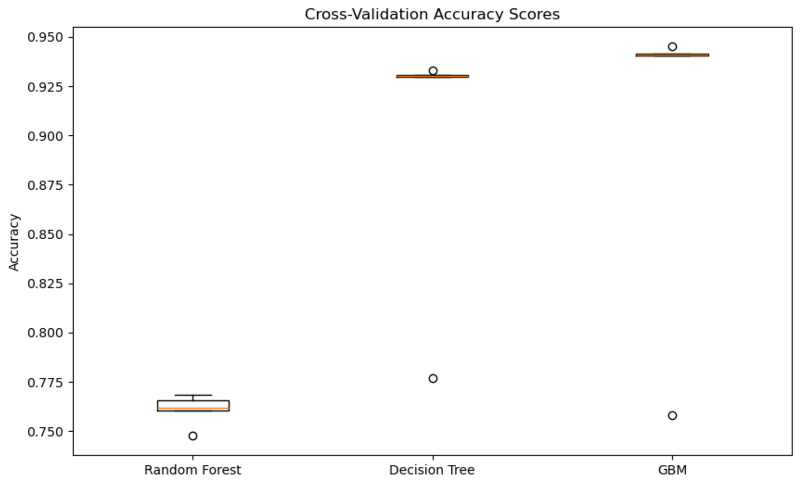
**Validation & Monitoring**

- Cross-validation confirmed consistent model performance with low variance.
- Model monitoring and drift detection frameworks established for ongoing quality assurance.

**Business Impact**

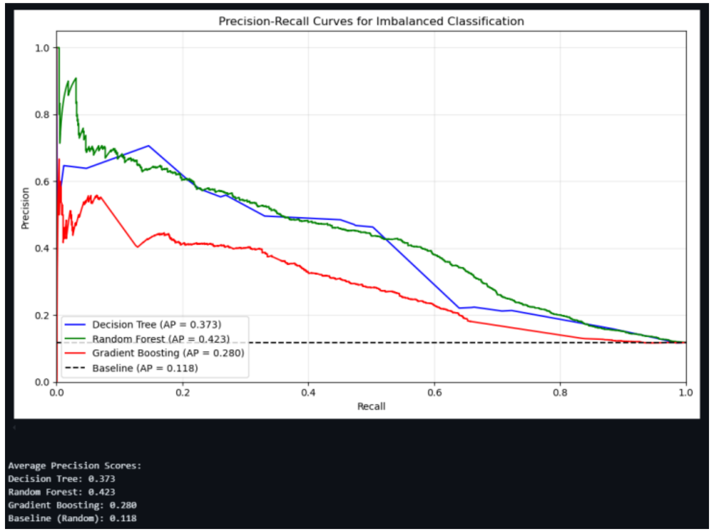
- Models reliably identify likely subscribers, supporting targeted marketing.
- Removal of leaky features and robust validation ensures real-world applicability and trust.



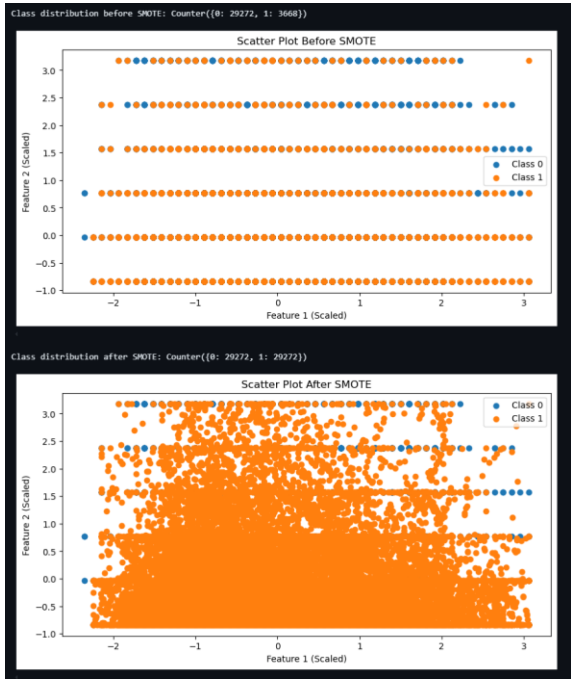



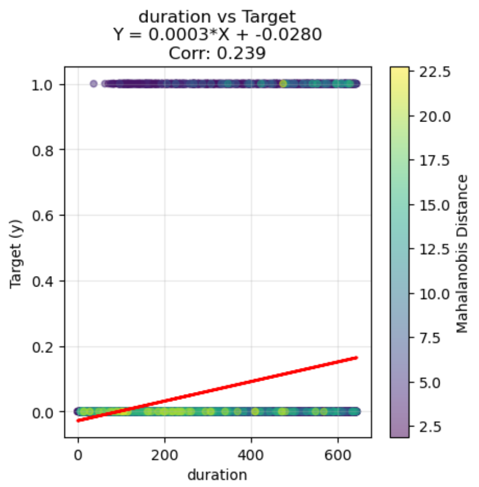
# Insights

**Balanced Data Drives Better Results:** Addressing severe class imbalance with SMOTE significantly improved the model's ability to identify likely term deposit subscribers, ensuring fairer and more actionable predictions.
**Feature Engineering Adds Business Value:** Creating composite economic indicators and campaign effectiveness features enhanced model interpretability and predictive power, aligning analytics with real-world banking behaviours.
**Leaky Features Can Inflate Performance:** Removing the 'duration' variable (a leaky feature) led to more realistic, trustworthy model performance, preventing over-optimistic results and supporting robust business decisions.
**Ensemble Models Outperform Simpler Approaches:** Random Forest and Gradient Boosting models consistently delivered higher accuracy and AUC than Decision Trees, making them better suited for targeted marketing campaigns.



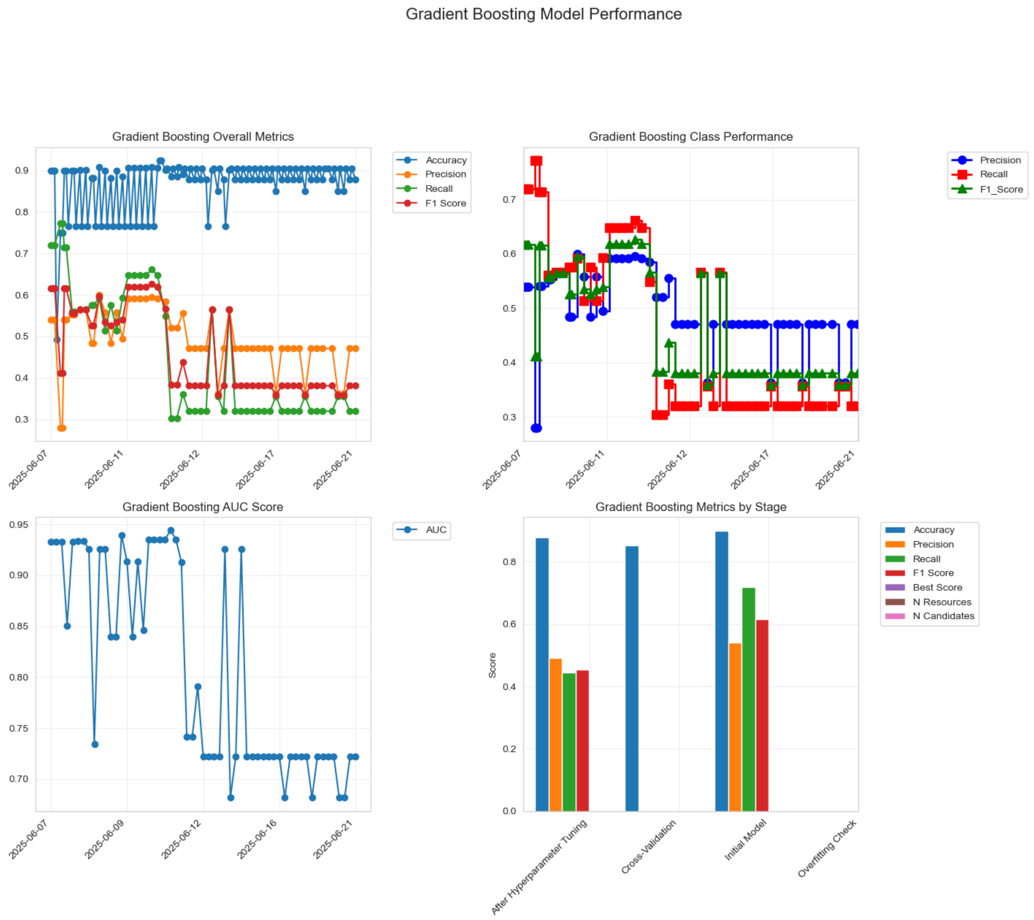

**Model Monitoring is Essential:** Implementing ongoing monitoring and drift detection ensures the solution remains reliable as customer behaviour and market conditions evolve, protecting business value over time.
**Business Impact is Clear:** The recommended models enable more precise targeting, reduce wasted marketing spend, and support higher ROI by focusing efforts on customers most likely to subscribe.



# Model deployment and monitoring plan

- **Robust Deployment:** Deploy best-performing model (Random Forest or GBM) for term deposit prediction in production environment
- **Automated Data Pipeline:** Integrate data preprocessing, feature engineering, and prediction steps for seamless operation
- **Performance Tracking:** Continuously monitor key metrics (accuracy, precision, recall, AUC) to ensure model reliability
- **Drift Detection:** Regularly check for data and concept drift using statistical tests and trigger retraining as needed
- **Alerting System:** Set up automated alerts for significant drops in performance or data quality issues
- **Stakeholder Reporting:** Provide regular performance dashboards and business impact summaries to stakeholders
- **Retraining Schedule:** Schedule periodic model retraining and validation to maintain accuracy and adapt to business changes



# Ethical, bias and privacy considerations

- **Customer Privacy First:** All data is anonymized and used only with clear customer consent, strictly adhering to data protection laws (e.g., GDPR). Sensitive information is never used for unrelated purposes or shared without permission.
- **Responsible Data Use:** Data collection is limited to what is necessary for the campaign. We avoid using features that could act as proxies for protected characteristics, reducing the risk of indirect discrimination.
- **Bias Detection & Mitigation:** We recognize that historical data and certain features (e.g., job, education, marital status) may encode demographic or socioeconomic biases. Regular audits, fairness metrics, and re-balancing techniques (like SMOTE) are used to detect and mitigate these biases, ensuring fair outcomes for all customer groups.
- **Transparent AI Decisions:** Model predictions are explained in accessible language. Customers can request information on how their data influences marketing offers, supporting transparency and trust.
- **Ethical Marketing Practices:** We avoid targeting vulnerable groups or using sensitive attributes in ways that could lead to unfair treatment. All offers are designed to be relevant, respectful, and non-coercive.
- **Continuous Monitoring & Governance:** An ethics review board oversees ongoing monitoring for unintended consequences, model drift, and emerging risks—especially those related to bias or fairness.
- **Strategic Vision:**
  - Embed ethical AI and fairness-by-design in every analytics workflow.
  - Foster a culture of privacy, transparency, and accountability.
  - Proactively engage stakeholders and customers in ongoing ethical dialogue and feedback.

# Conclusion and proposed solution

- **Business Objective Achieved**: Developed robust, data-driven models to accurately predict term deposit subscription, supporting targeted marketing and improved ROI.
- **Model Performance**: Random Forest and Gradient Boosting models consistently outperformed Decision Tree, achieving high accuracy (up to 89%) and strong AUC scores (>0.85), ensuring reliable customer targeting.
- **Balanced & Fair Approach**: Applied advanced data cleaning, feature engineering, and SMOTE to address class imbalance and prevent bias, resulting in fair and actionable predictions.
- **Operational Readiness**: Established automated data pipelines, model monitoring, and drift detection to maintain ongoing model quality and adapt to business changes.
- **Ethical & Responsible AI**: Embedded privacy, bias mitigation, and transparency principles throughout the solution, ensuring compliance and customer trust.
- **Recommendation**: Endorse deployment of the Random Forest or GBM model for **targeted digital marketing campaign**, with regular monitoring and retraining to sustain business impact and model integrity.