Portuguese Bank Marketing Project

PROBLEM OF STATEMENT:

- Prepare a complete data analysis report on the given data.
- Creating a predictive model which will help the bank marketing team to know which customer will buy the product.
- Suggestions to the Bank market team to make customers buy the product**.**

Description:

Input Variables:

1. Age: Numeric variable representing the age of the client.
2. Job: Categorical variable indicating the type of job. Categories include admin, blue-collar, entrepreneur, housemaid, management, retired, self-employed, services, student, technician, unemployed, and unknown.
3. Marital: Marital status of the client. Categories include divorced, married, single, and unknown (which can also represent widowed clients).
4. Education: Categorical variable describing the client's education level. Categories include basic education levels (4y, 6y, 9y), high school, illiterate, professional course, university degree, and unknown.
5. Default: Indicates if the client has credit in default. Categories are no, yes, and unknown.
6. Housing: Indicates if the client has a housing loan. Categories are no, yes, and unknown.
7. Loan: Indicates if the client has a personal loan. Categories are no, yes, and unknown.

Attributes Related to the Last Contact of the Current Campaign:

8. Contact: Communication type used to contact the client. Categories include cellular and telephone.
9. Month: Last contact month of the year. Categories include January to December.
10. Day_of_week: Last contact day of the week. Categories include Monday to Friday.
11. Duration: Duration of the last contact in seconds. Note: This attribute strongly influences the target, and its value is only known after the call. It's recommended for benchmarking purposes and can be omitted from a predictive model.

Other Attributes:

12. Campaign: Number of contacts performed during this campaign for the client, including the last contact.
13. Pdays: Number of days passed after the client was last contacted from a previous campaign. 999 represents clients not contacted previously.
14. Previous: Number of contacts performed before this campaign for the client.
15. Poutcome: Outcome of the previous marketing campaign. Categories are failure, nonexistent, and success.

Social and Economic Context Attributes:

16. Emp.var.rate: Quarterly employment variation rate indicator (numeric).
17. Cons.price.idx: Monthly consumer price index indicator (numeric).
18. Cons.conf.idx: Monthly consumer confidence index indicator (numeric).

19. Euribor3m: Daily euribor 3 month rate indicator (numeric).

20. Nr.employed: Quarterly number of employees indicator (numeric).

21. Output Variable (Target):

22. Y: Binary variable indicating whether the client subscribed to a term deposit. Categories are yes and no.

Model Building Report:

1. Logistic Regression Model:

Steps Taken:

- Imported necessary libraries and loaded the dataset.
- Applied SMOTE (Synthetic Minority Over-sampling Technique) to address class imbalance.
- Created a Logistic Regression model and trained it on the resampled data.
- Evaluated the model's performance using precision, recall, and f1-score.

Hyperparameter Tuning Results:
Best Hyperparameters:
Solver: 'liblinear'
Penalty: 'l1'
Max Iter: 300
C: 1

Performance Metrics:
Precision Score: 0.91
Recall Score: 0.99
F1 Score: 0.95

2. Random Forest Model Report

1. Steps Taken:

- Imported necessary libraries and loaded the dataset.
- Applied SMOTE (Synthetic Minority Over-sampling Technique) to address class imbalance.
- Built a Random Forest model and trained it on the resampled data.
- Evaluated the model's performance on both the training and testing sets using key metrics: accuracy, precision, recall, and F1 score.
- Performed hyperparameter tuning using GridSearchCV to find the best combination of hyperparameters for the Random Forest model.

2. Hyperparameter Tuning Results:

Best Hyperparameters:
n_estimators: 500
min_samples_split: 2
min_samples_leaf: 4
max_features: 'log2'
max_depth: 50
bootstrap: False

3. Performance Metrics:

Train Accuracy: 1.00
Train F1 Score: 1.00
Test Accuracy: 0.86
Test Precision: 0.95
Test Recall: 0.90
Test F1 Score: 0.92

3. Decision Tree Model Report

1. Steps Taken:

- Imported necessary libraries and loaded the dataset.
- Applied SMOTE (Synthetic Minority Over-sampling Technique) to address class imbalance.
- Built a Decision Tree model and trained it using the resampled data.
- Evaluated the model's performance using key metrics: accuracy, precision, recall, and F1 score.
- Performed hyperparameter tuning to find the optimal configuration using GridSearchCV.

2. Hyperparameter Tuning Results:

Best Hyperparameters:
Criterion: 'gini'
Max Depth: 5
Max Features: None
Min Samples Leaf: 4
Min Samples Split: 2

3. Performance Metrics:

Train Accuracy: 1.00
Train F1 Score: 1.00
Test Accuracy: 0.86
Test Precision: 0.95
Test Recall: 0.90
Test F1 Score: 0.92

4. Gradient Boosting Model Report

1. Steps Taken:

- Imported necessary libraries and loaded the dataset.

- Applied SMOTE (Synthetic Minority Over-sampling Technique) to balance the dataset and mitigate class imbalance.
- Built a Gradient Boosting model using default settings and trained it on the resampled data.
- Evaluated the model's performance on both the training and testing sets using key metrics: accuracy, precision, recall, and F1 score.
- Performed hyperparameter tuning using GridSearchCV to find the best combination of hyperparameters.

2. Hyperparameter Tuning Results:

Best Hyperparameters:
n_estimators: 50
max_depth: 5
learning_rate: 0.1

3. Performance Metrics:

Train Accuracy: 0.92
Train F1 Score: 0.96
Test Accuracy: 0.91
Test Precision: 0.93
Test Recall: 0.97
Test F1 Score: 0.95

Best Performing Model:

- Highest Test Accuracy (0.91): Ensures strong predictive performance on unseen data.
- Best F1 Score (0.95): Balances precision and recall, reducing false positives and false negatives.
- Better Generalization: Unlike Decision Tree and Random Forest, Gradient Boosting does not overfit.
- Strong Recall (0.97): Effectively identifies positive cases, making it ideal for banking applications.

The Gradient Boosting model stood out as the most effective, providing the best balance between accuracy, precision, recall, and F1-score. Unlike Decision Trees and Random Forest, which showed signs of overfitting, Gradient Boosting generalized well to unseen data. Its high recall makes it particularly useful in banking applications where correctly identifying potential customers is crucial.

While Logistic Regression was simpler and interpretable, it lacked the predictive power needed for complex patterns. Overall, Gradient Boosting emerges as the optimal model for this project, offering strong predictive capability and reliability.

Suggestions to the Bank market team to make customers buy the product:

1. Targeted Offers

Targeted offers involve creating promotions or discounts specifically designed to appeal to certain groups of customers. The goal is to incentivize purchases by providing deals that align with the customers' preferences and needs.

- Customer Profiling: Create detailed profiles of customers based on demographics, purchase history, and preferences.
- Segment-Specific Offers: Tailor offers to different customer segments. For example, offer young professionals deals on technology products.
- Past Purchase Behavior: Analyze previous purchases and suggest related or complementary products.
- Exclusive Discounts: Provide early access or special discounts to specific customer groups to create a sense of exclusivity.
- Limited-Time Offers: Create urgency by offering time-sensitive deals to encourage quick decisions.
- Personalized Promo Codes: Generate unique promo codes for each customer to enhance engagement.
- Feedback-Driven Offers: Use customer feedback to craft offers that address their specific interests or concerns.

---

2. Timing in Marketing

Timing in marketing refers to strategically choosing when to deliver messages, content, and promotions to maximize engagement and conversions.

- Analyze Audience Behavior: Identify when your audience is most active.
- Use Data and Analytics: Leverage past campaign data to determine optimal engagement times.
- Social Media Insights: Use platform analytics to post when followers are most active.
- Geographical Considerations: Adjust timing for different time zones if your audience is global.
- Event-Based Timing: Align campaigns with holidays, events, or seasonal trends.
- Behavioral Triggers: Automate messages based on customer actions, such as abandoned carts.
- Frequency and Consistency: Maintain a regular posting and communication schedule.
- Testing and Optimization: Use A/B testing to refine timing strategies.

---

3. Exclusive Discounts in Marketing

Exclusive discounts create a sense of privilege and urgency, encouraging customer engagement.

- Segmentation: Identify key customer groups for exclusive discounts, such as frequent buyers or subscribers.
- Personalization: Tailor offers based on customer preferences and past purchases.
- Promotion: Share discounts through multiple channels like email, SMS, and social media.
- Clear Communication: Emphasize exclusivity and the benefits of the offer.

- Redemption Process: Provide clear instructions for redeeming the discount, such as promo codes or unique links.
- Expiry Date: Create urgency with a defined expiration date.
- Measure and Analyze: Track campaign performance to understand customer response.

---

4. Multi-Channel Approach in Marketing

A multi-channel marketing approach uses different platforms to engage with customers across multiple touchpoints.

- Diverse Channels: Utilize email, social media, SMS, websites, and in-store promotions.
- Consistent Messaging: Maintain uniform branding and messaging across all platforms.
- Customer-Centricity: Customize messages based on the preferences of each audience segment.
- Cross-Channel Integration: Ensure that various channels work together seamlessly.
- Timing Consideration: Optimize posting and engagement times for each channel.
- Data Utilization: Analyze customer interactions to refine marketing strategies.

---

5. Simplifying the Process in Marketing

Simplifying the customer journey makes interactions with your brand more seamless and user-friendly.

- Improved User Experience: A smooth experience increases customer satisfaction.

- Higher Conversions: Reducing complexity helps customers complete purchases or sign-ups.
- Reduced Abandonment: Simplifying checkout and forms lowers cart abandonment rates.
- Enhanced Brand Perception: Brands that prioritize simplicity appear more modern and customer-focused.

By streamlining processes, removing unnecessary steps, and making interactions intuitive, businesses can improve engagement, satisfaction, and overall conversion rates.