

## **Summary Report**

### **Introduction**

The objective of this assignment was to analyze factors influencing lead conversion for X Education, develop strategies to optimize lead conversion efforts, and create a detailed presentation of the findings. The primary goals were to identify key variables affecting lead conversion, determine the most important categorical/dummy variables, and recommend business strategies for different operational phases.

### **Data Preprocessing**

The first step involved data preprocessing to clean and prepare the data for modeling. Missing values were filled with appropriate defaults, such as mean values for numerical columns, and categorical variables were converted into dummy variables using one-hot encoding. This ensured the dataset was complete and suitable for model training.

### **Model Selection**

A RandomForestClassifier was selected for the analysis due to its robustness and ability to handle a mix of numerical and categorical data. The dataset was split into training and testing sets to evaluate the model's performance. The model was trained on the training set and tested on the testing set to ensure accuracy and generalizability.

### **Feature Importance**

Feature importance was extracted from the RandomForestClassifier model to identify the variables most significantly impacting lead conversion. The top three variables contributing the most towards lead conversion were identified, along with the top three categorical/dummy variables. This step was crucial in understanding which factors the business should focus on to enhance lead conversion rates.

## **Results and Business Implications**

The analysis revealed key variables influencing lead conversion, providing valuable insights for X Education. The top three variables were:

1. **Total Time Spent on Website**
2. **Lead Source\_Welingak Website**
3. **Lead Origin\_Landing Page Submission**

Similarly, the top three categorical/dummy variables were identified:

1. **Lead Source\_Welingak Website**
2. **Lead Origin\_Landing Page Submission**
3. **Specialization\_Select**

These findings translated into actionable business strategies:

1. **For aggressive lead conversion during intern phases:**
  - Prioritize high-probability leads.
  - Segment and allocate leads among interns to ensure focus on the most promising leads.
  - Continuously monitor and adjust lead conversion rates to optimize performance.
2. **To minimize unnecessary phone calls after meeting targets:**
  - Set a higher threshold for lead scores to ensure only the most promising leads are contacted.
  - Use automated filtering to identify and remove lower probability leads.

- Allocate the sales team to focus on other important tasks once targets are met.

## Visualizations

Two key visualizations were created to support the findings:

1. **Feature Importance Plot:** A bar plot showing the top features contributing to lead conversion, providing a visual representation of the most significant variables.
2. **ROC Curve:** A plot displaying the Receiver Operating Characteristic (ROC) curve, illustrating the model's performance in distinguishing between converted and non-converted leads.

## Summary and Key Takeaways

The analysis provided valuable insights into the factors driving lead conversion for X Education. The identification of key variables and categorical/dummy variables offered clear areas for the business to focus on to improve conversion rates. The recommended strategies for different operational phases were based on the model's findings and aimed at optimizing sales efforts. The visualizations effectively communicated the results, making it easier for stakeholders to understand and act on the insights.

## Learnings

Through this assignment, several important learnings were gathered:

1. **Importance of Data Preprocessing:** Ensuring data completeness and accuracy is crucial for reliable model performance.
2. **Model Selection and Evaluation:** Choosing the right model and properly evaluating its performance can significantly impact the quality of insights derived.

3. **Feature Importance:** Understanding which variables most influence the target outcome can guide strategic business decisions.
4. **Visualization:** Effective visualizations are essential for communicating complex findings in an accessible manner.
5. **Strategic Recommendations:** Translating analytical findings into actionable business strategies is a key step in delivering value from data analysis.