**LEAD SCORE MODEL IMPLEMENTATION**

SUMMARY

**By Prashant Kumar Singh**

The implemented lead scoring model is a robust solution designed to identify potential hot leads efficiently. Leveraging Python, pandas, and scikit-learn, the model underwent a comprehensive data cleaning and preprocessing phase. Missing values were addressed, skewed categorical columns were treated, and outlier handling techniques were applied, ensuring the dataset's quality and reliability. Key features such as Lead Source, City, and Specialization were systematically analysed and processed to improve the model's predictive capabilities.

The data preparation phase involved crucial steps like train-test splitting, scaling of numerical features, and creating dummy variables for categorical columns. The feature selection process utilized Recursive Feature Elimination (RFE) with logistic regression, refining the model by eliminating features with high p-values and Variance Inflation Factors (VIF). This meticulous feature engineering contributed to a more efficient and accurate lead scoring model.

The logistic regression model demonstrated commendable performance on the training dataset, achieving a high sensitivity of 79.99%. Sensitivity is a crucial metric for a lead scoring model, indicating its effectiveness in correctly identifying potential hot leads. The model's robustness was further validated through precision-recall trade-off analysis and Receiver Operating Characteristic (ROC) curve evaluation.

The model's optimal threshold was carefully selected to strike a balance between sensitivity and accuracy, crucial for effective lead prioritization. The implementation process ensures consistency in preprocessing steps between the training and test datasets, making the model ready for deployment. The final model is poised to streamline the sales process by efficiently identifying leads with a high likelihood of conversion, empowering sales teams with valuable insights and improving overall conversion rates. Overall, the lead scoring model represents a data-driven approach to enhance lead management and contribute to the overall success of sales and marketing efforts.