

DOMAIN ORIENTED CASE STUDY

By
Anand
Anjali
Aniketh

- **Objective**

- Briefly introduce the problem statement and objectives of the case study.
- State the relevance of the analysis and its potential impact.

- ▶ **Slide 2: Data Understanding and Preparation**

- **Data Overview**

- Summary of the dataset structure (columns, rows).
- Mention any initial challenges or observations during data loading.

- **Data Cleaning**

- Discuss steps taken to handle missing values.
- Highlight any data transformations (converting data types, standardizing formats).

- **Data Quality Checks**

- Examples of checks performed (e.g., missing values, outliers).
- How data integrity was maintained for accurate analysis

► Slide 3: Exploratory Data Analysis (EDA)

- **Overview**

- Purpose and significance of EDA in understanding the dataset.
- Mention tools and techniques used (e.g., visualizations, summary statistics).

- **Key Findings**

- Present insights derived from EDA (e.g., distribution of target variable, correlations).
- Highlight any surprising or notable discoveries that influenced subsequent steps.

- **Feature Importance**

- Discuss preliminary insights into feature importance.
- Visualize or list top features identified through initial analysis.

► Slide 4: Feature Engineering

- **Purpose of Feature Engineering**

- Explain why feature engineering is crucial for improving model performance.
- Outline specific techniques used (e.g., creating interaction terms, transforming variables).

- **New Features**

- Provide examples of newly created features and their rationale.
- Illustrate how these features enhance the predictive power of the models.

► Model Development

- **Model Selection**

- List the variety of models explored (e.g., Random Forest, Logistic Regression).
- Mention the rationale for choosing these models based on the problem requirements.

- **Training Process**

- Overview of the training process (e.g., data splitting, cross-validation).
- Emphasize the need for models to output probability scores for reliable predictions.

► Slide 6: Model Selection and Hyperparameter Tuning

- **Handling Class Imbalance**

- Discuss techniques used to address class imbalance (e.g., SMOTE).
- Explain why handling imbalance is critical for accurate model performance.

- **Hyperparameter Tuning**

- Describe the approach to tuning model hyperparameters (e.g., GridSearchCV).
- Highlight specific parameters tuned and their impact on model performance.

► Model Evaluation

- **Evaluation Metrics**

- Present the evaluation metrics used (e.g., ROC AUC, accuracy, precision-recall).
- Justify why these metrics were chosen based on the problem context.

- **Results and Comparison**

- Display results of model evaluation (e.g., performance metrics for each model).
- Compare the effectiveness of different models and their suitability for deployment.

► Slide 8: Business Impact and Recommendations

- **Key Insights**

- Summarize important findings from the analysis (e.g., critical features, model predictions).
- Connect these insights to potential business implications.

- **Recommendations**

- Provide actionable recommendations based on data-driven insights.
- Discuss how stakeholders can leverage these recommendations for decision-making.

► Conclusion

- **Summary of Findings**

- Recapitulate the main findings and outcomes of the case study.
- Reaffirm the relevance and impact of the analysis for the business or domain.

