

LEAD SCORING CASE STUDY SUMMARY REPORT

Introduction

The lead scoring case study aimed to develop a predictive model for a business to identify and prioritize potential leads based on their likelihood to convert into customers. The assignment involved several key steps, including data pre-processing, feature engineering, model selection, and evaluation. This report outlines the methodology, findings, and key takeaways from the lead scoring case study.

Data Pre-processing

The first step in the assignment was data pre-processing. This involved cleaning the dataset, handling missing values, and encoding categorical variables. Data pre-processing is a critical step as it ensures the data used for modelling is accurate and complete.

Feature Engineering

Feature engineering played a significant role in improving model performance. We created new features based on domain knowledge and business understanding, such as lead source, lead activity, and lead engagement. Feature scaling and normalization were also applied to make the features more suitable for modelling.

Exploratory Data Analysis (EDA)

EDA was conducted to gain insights into the dataset. Visualization techniques were used to identify trends, correlations, and patterns within the data. EDA helped in understanding which features were most influential in lead conversion and provided guidance for feature selection.

Model Selection

Several machine learning models were considered for lead scoring. Logistic Regression, Random Forest, and Gradient Boosting were among the models evaluated. Cross-validation was used to assess model performance and select the most suitable one.

Model Training and Evaluation

The selected model was trained on the dataset and evaluated using appropriate metrics. Metrics such as accuracy, precision, recall, F1-score, and ROC AUC were used to assess model performance. Additionally, a confusion matrix was generated to understand the model's ability to correctly classify leads.

Hyper parameter Tuning

Hyper parameter tuning was performed to optimize the selected model's performance. Grid search and random search techniques were used to find the best combination of hyper parameters. This step helped improve the model's predictive accuracy.

Model Deployment and Automation

To make the lead scoring process more efficient, the final model was deployed for automation. This involved creating scripts or pipelines to score new leads automatically as they entered the system. This deployment allowed the business to prioritize and nurture leads effectively.

Key Learnings

1. **Data Quality is Vital:** The quality of data significantly impacts the performance of the lead scoring model. Data pre-processing and cleaning are critical steps to ensure accurate predictions.
2. **Feature Engineering Matters:** Creating relevant features based on domain knowledge can enhance model performance. Understanding the business context is crucial for feature engineering.
3. **Model Selection is a Trade-Off:** Choosing the right model involves a trade-off between interpretability and predictive power. Decision-makers must consider the business's needs and goals.
4. **Evaluation Metrics Matter:** The choice of evaluation metrics depends on the business's objectives. Precision and recall may be more important than accuracy in lead scoring.
5. **Hyper parameter Tuning is Necessary:** Fine-tuning hyper parameters can significantly improve model performance. It's a critical step in model development.
6. **Deployment is Key for Real-World Impact:** Deploying the model for automation is essential for realizing the benefits of lead scoring. It streamlines lead management and improves efficiency.

Conclusion

In conclusion, the lead scoring case study provided valuable insights into the process of identifying and prioritizing potential leads. The assignment involved data pre-processing, feature engineering, model selection, and evaluation. Key learnings included the importance of data quality, feature engineering, and model selection based on business needs. Additionally, deploying the model for automation was crucial for real-world impact. This case study demonstrated the practical application of machine learning in lead management and customer acquisition.