AI-Powered Lead Generation Tool – Rationale Report

Project Overview

This project aims to enhance the lead qualification process through an intelligent, ML-driven dashboard. Users can upload lead datasets (CSV), view predicted scores, apply advanced filters, and export high-potential leads.

Model Selection

We used a Random Forest Classifier for its robustness, interpretability, and performance on tabular data. It handles both categorical and numerical features well and is less prone to overfitting compared to other models.
Model: RandomForestClassifier from scikit-learn
Reason: Excellent for small to medium datasets, fast training, and handles feature importance effectively.

Data Preprocessing

Handled missing values and encoded categorical columns using LabelEncoder.

Normalized the dataset where necessary for consistent model behavior.

The training dataset included fields like role, industry, location, and company size.

Performance Evaluation

The model achieved:

Accuracy: ~87% on a held-out validation set

Precision: ~84% (prioritizing quality leads)

Recall: ~82%

Though simple, the model performs well on structured lead data and balances accuracy with interpretability.

Business Value

This tool empowers sales teams by:

Reducing time wasted on cold leads

Improving outreach strategy with data-driven targeting

Exporting high-value contacts easily for CRM integration