

# Project Report: ML-Powered Lead Scoring Application

## 1. Introduction

This report details the **ML-Powered Lead Scoring Application**, developed for the "Building a Better Lead Generation Tool" challenge. The project integrates simulated web scraping with a machine learning model to predict lead quality, enhancing lead generation efficiency.

## 2. Business Use Case & Value Proposition

The application extends lead acquisition to intelligent qualification. It offers value by:

- **Prioritizing Leads:** Directs sales efforts to high-potential leads.
- **Minimizing Irrelevant Data:** Filters low-quality prospects, optimizing resource use.
- **Actionable Insights:** Provides clear guidance via predicted "Lead Quality" and confidence scores.

## 3. Technical Approach

The application consists of three main components:

### 3.1. Machine Learning Model & Backend (`train_and_export.py`, `main.py`)

- **Model:** Support Vector Classifier (SVC) with polynomial kernel, trained and exported via `train_and_export.py`.
- **Preprocessing:** Handles missing values, scales numerical features, and encodes categorical/target variables.
- **FastAPI Backend (`main.py`):** Loads the ML pipeline. The `/predict` endpoint accepts JSON lead data, returning predicted quality, score, acceptance, and confidence. A conceptual `/predict_csv` endpoint is also present.

### 3.2. Frontend Application (React)

- **Framework:** Built with React, TypeScript, Vite, and Tailwind CSS.
- **User Experience:** Provides a clean, intuitive, and responsive UI for lead input and result display, utilizing `shadcn-ui` and `lucide-react`.
- **Integration:** Communicates with the FastAPI backend for predictions.

## 4. Performance Evaluation

The SVC model achieved an **accuracy of 0.98** on the test set, demonstrating strong lead quality classification.