1. Problem Statement

1.1 Introduction

Lead scoring is a critical aspect of B2B sales, allowing businesses to prioritize potential customers based on their likelihood of conversion. Traditional lead scoring methods rely heavily on manual assessment and rule-based systems, which can be inefficient and inaccurate. The logistics industry, in particular, deals with large volumes of leads, making automated and intelligent lead scoring essential.

1.2 Key Challenges

- Inefficient manual lead qualification processes
- Lack of predictive insights for lead prioritization
- Difficulty in understanding the key factors influencing lead conversion
- Limited integration of AI-driven insights with CRM systems

1.3 Project Goals

This project aims to develop an AI-based lead scoring system tailored for B2B logistics companies. The model should:
- Utilize machine learning to score leads based on historical data
- Provide explainability for the scores through SHAP (SHapley Additive exPlanations)
- Incorporate voice AI for user interaction and feedback collection
- Allow continuous learning and improvement through human feedback

2. Approach Taken

2.1 Data Collection and Preprocessing

Given the absence of a real-world dataset, a synthetic database was generated using ChatGPT. This dataset included features such as:
- Company size
- Industry type
- Previous interactions
- Website activity
- Engagement with marketing campaigns
- Lead source and referral data

The dataset was preprocessed with the following steps:

- Handling missing values: Imputed where necessary
- Feature engineering: Created derived features for better predictive performance
- Normalization and encoding: Ensured compatibility with machine learning models

2.2 Model Development

The lead scoring model was trained using machine learning algorithms, including:
- Random Forest (for robust decision-making)
- XGBoost (for boosting performance)
- Logistic Regression (for baseline comparison)

2.3 Explainability with SHAP

Since this was the first time using SHAP, it was employed to interpret the model's decision-making process. SHAP values provided insights into which factors contributed most to a lead's score, helping sales teams understand why a lead was prioritized.

2.4 API and Dashboard Development

- A RESTful API was developed to integrate lead scoring into existing CRM systems.
- An interactive dashboard using Streamlit was created to visualize lead scores and SHAP explanations.
- A voice interaction module was implemented to explain scores and collect verbal feedback.

3. Challenges Faced

3.1 Working with Synthetic Data

- Challenge: The synthetic dataset, while useful, lacked real-world unpredictability and biases.
- Solution: Introduced noise and variability to simulate realistic data patterns.

3.2 First-Time Use of SHAP

- Challenge: Understanding SHAP and effectively implementing it in the model was challenging.
- Solution: Studied SHAP documentation and tested various visualizations to ensure meaningful insights.

3.3 Model Overfitting

- Challenge: Initial models performed exceptionally well on training data but less effectively on unseen data.
- Solution: Applied regularization techniques and performed hyperparameter tuning.

3.4 Voice AI Integration

- Challenge: Converting lead insights into natural language explanations.
- Solution: Used text-to-speech models and predefined templates for structured responses.

4. Future Improvements

4.1 Enhancing Data Quality

- Incorporating real-world lead data to replace synthetic data for better model generalization.

4.2 Advanced Model Selection

- Experimenting with deep learning models for even better lead prediction.

4.3 Improved Explainability

- Enhancing SHAP-based visualizations for non-technical users.

4.4 Expanding Voice AI Capabilities

- Making the voice assistant more interactive with natural conversation capabilities.

4.5 Continuous Learning

- Implementing a reinforcement learning mechanism where human feedback improves future predictions.

This report outlines the development, challenges, and future prospects of the AI-driven lead scoring system. The project demonstrates how AI can revolutionize B2B sales by providing intelligent, explainable, and interactive lead qualification solutions.