

Two vertical lines, one black and one green, are positioned in the top left corner of the page.

# **CASE STUDY APPLICATION**

PART 2

A horizontal line with a green segment on the left and a black segment on the right, located below the text 'PART 2'.Abstract geometric shapes in the bottom left corner, including a black triangle, a grey triangle, and a green triangle.Three large, light green, wavy, curved lines that sweep across the bottom right portion of the page.

# PROBLEM SCOPE

**Problem:**

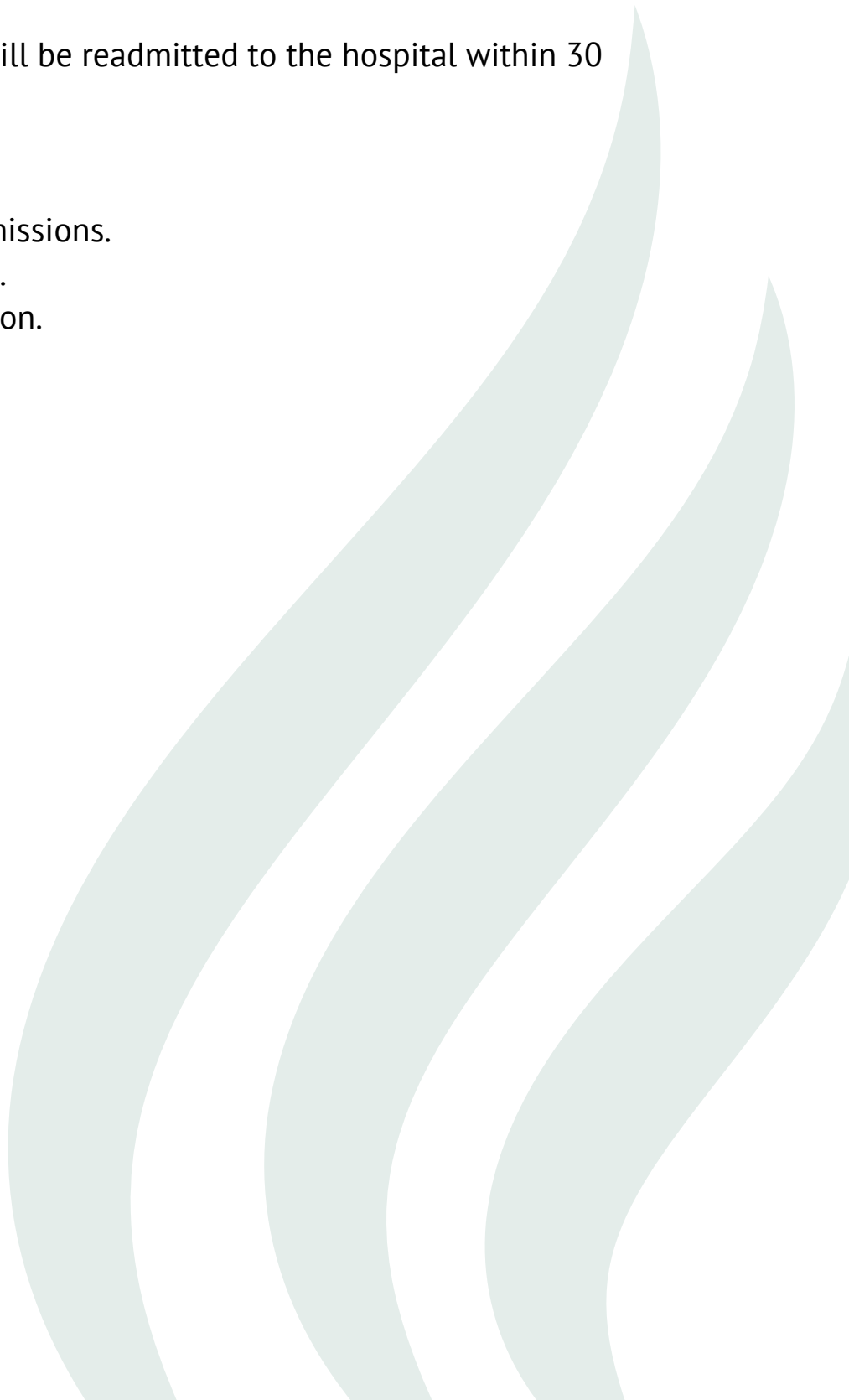
Predict whether a patient will be readmitted to the hospital within 30 days of discharge.

**Objectives:**

- Reduce preventable readmissions.
- Improve patient outcomes.
- Optimize resource allocation.

**Stakeholders:**

- Hospital administrators
- Clinicians and care teams
- Patients
- Data analysts/IT staff



# DATA STRATEGY

## Data Sources:

- Electronic Health Records (EHRs): diagnoses, procedures, medications, lab results, discharge summaries
- Demographics: age, gender, ethnicity
- Admission/discharge details
- Comorbidities (e.g., diabetes, hypertension)

## Ethical Concerns:

- Patient Privacy: Protect sensitive health information; ensure data is anonymized and securely stored.
- Bias & Fairness: Avoid models that disadvantage certain groups (e.g., by gender, ethnicity).

## Preprocessing Pipeline:

### 1. Missing Value Handling:

- Identify and impute or drop missing data.

### 2. Feature Engineering:

- Extract systolic/diastolic blood pressure from text fields.
- Encode binary features (e.g., diabetes, hypertension) as 0/1.
- One-hot encode multi-category features (e.g., gender, discharge destination).

### 3. Scaling/Normalization:

- Apply if needed for certain models (e.g., SVMs, neural networks).

### 4. Train/Test Split:

- Use an 80/20 split for model validation.