Assignment: Build a Classifier to Predict Injury Incidents

**Objective:**

You are tasked with building a machine learning classifier that can predict whether a safety incident report involves an injury based on the text of the report. This is a critical task for enhancing safety processes at the airline by allowing quicker identification of incidents that require immediate attention.

**Data:**

You have been provided with a dataset (safety\_incident\_reports.csv) containing 2000 safety incident reports from airline employees. The dataset includes the following columns:

* **Description**: The text description of the safety incident, which can be in English or Dutch. The description may or may not indicate that an injury occurred.
* **label**: A binary label where:
  + 1 indicates the incident resulted in an injury.
  + 0 indicates no injury occurred.

**Goal:**

Your goal is to create a machine learning model that can predict the IsInjury column using the text in the FormDescription. The model should classify whether the report describes an injury-related incident (1) or not (0). The code should be in Python. Here are some guidelines:

1. Data Understanding and Preprocessing

2. Model Selection and Training

3. Model Evaluation

**Deliverables:**

A Jupyter Notebook (or Python script) that includes:

* Data loading and preprocessing steps.
* Feature engineering.
* Model training, evaluation, and optimization.
* Visualizations (e.g., confusion matrix, performance metrics).
* A brief report explaining:
* The modeling process.
* The evaluation of your classifier.
* Any challenges faced and how you overcame them.