

# ML Final Project

**Group members:** Dailin Li, Tyler Abele, Margaux Reynolds

**Course:** DS 4021

**Date:** 11/17/2025

**Motivation narrative:**

This project seeks to explore how patterns in personal habits—such as smoking, alcohol consumption, and exposure to daily stressors—as well as symptoms like coughing, wheezing, and fatigue, relate to an individual's likelihood of having lung cancer.

**Research Question:**

Which specific features are most predictive of lung cancer presence?

**Dataset description:**

The dataset used in this study originates from an online lung cancer prediction system aimed at helping individuals assess their cancer risk at low cost. It includes 309 records with attributes covering demographic information, personal habits, self-reported symptoms, and medical history indicators. Each record is labeled with a binary lung cancer diagnosis (YES/NO). The features include smoking behavior, yellow fingers, anxiety, chronic disease, fatigue, wheezing, alcohol consumption, coughing, chest pain, swallowing difficulty, and several other relevant symptoms.

**Github repo:** [https://github.com/dylanlidailin/ML\\_final\\_project](https://github.com/dylanlidailin/ML_final_project)

**References:** <https://www.kaggle.com/datasets/mysarahmadbhat/lung-cancer?resource=download&select=survey+lung+cancer.csv>