# Startup Failure Predictor

Elisa Holland

# Topic

Why do startups fail, and can we predict it?

Startups drive innovation and job creation, but many fail early due to poor market fit, limited funding, or competition. This project explores whether we can predict startup failure using modeling

# Objective

- Build a logistic regression model that can classify whether a startup is likely to fail or succeed based on features like funding, industry, and location.
- This helps investors and entrepreneurs better understand early-stage risk.

# Description of Data

**Sources**

- Failed Startups CSVs (multiple industries)
- Kaggle dataset of funded startups (assumed successful)

**Final Dataset**

- ~3,400 startups
- Combined failures and successes
- Label: 1 = failed, 0 = successful

**Features**

- Industry
- City
- Funding amount
- Years of operation

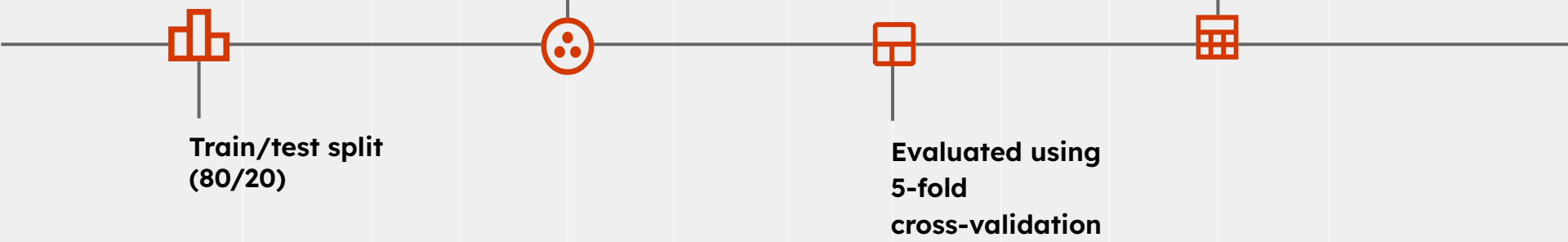
**Preprocessing**

- Standardized funding values
- One-hot encoded text fields
- Log-transformed funding
- Missing data handled by imputation or zero-filling

# Model: Logistic Regression with L2 Regularization

**Interpretable, Good baseline for binary classification, Works well with imbalanced data**

Methods:



# Results

## Performance on Test Set

Accuracy: **99.7%**

F1 Score: **1.00 (success)**, **0.99 (failure)**

AUC-ROC: **0.99**

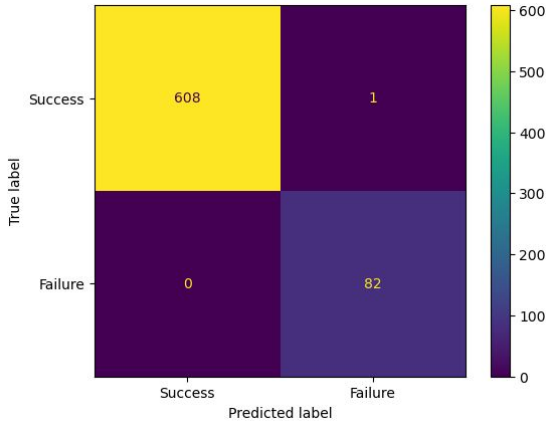
## Top Predictive Features

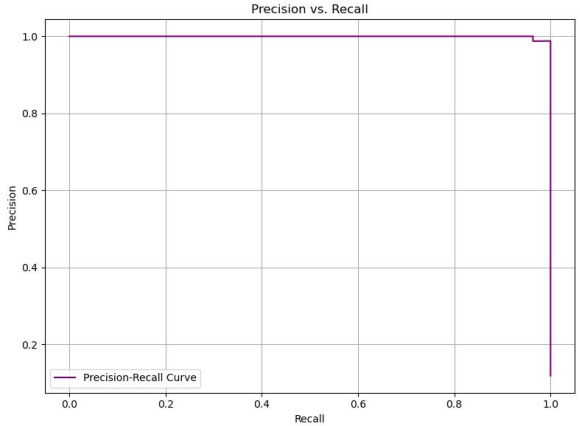
**1** Industry Type

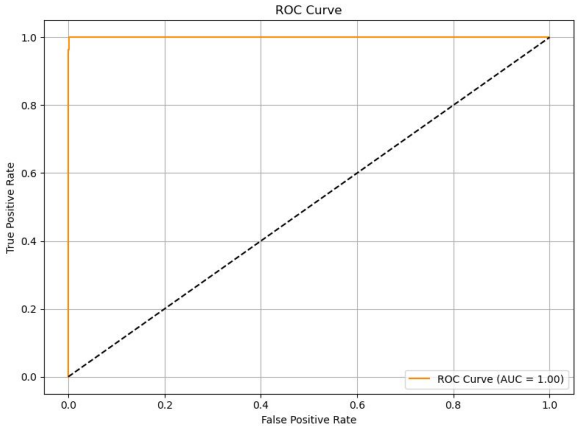
**2** City Location

**3** Funding Amount

## Visuals



# Thank You!