Cervical Cancer Risk Factors

Estela Perez, Angie Bolden, Ramiro Cervantes, Paola Moreno

Our purpose is to train an algorithm to recognize risk indicators and predict if a patient is at risk for Cervical Cancer.

Team Leader: (Ramiro)

Coordinating Github repository, presentation, and other project management tasks.

Front-End (Angie)

Create a site to display our visualizations, analysis, data sources, etc.

Host: AWS (Estela)

Our code will utilize Spark to retrieve our data and will be hosted through Amazon’s AWS. Our code will also clean, normalize, and standardize data prior to modeling.

Model: Logistical Regression (Paola)

Initialize, train, and evaluate our model. Measure our model’s predictive power by measuring classification accuracy and R-squared.

Source: Dataset was originally obtained from UCI Repository

<https://www.kaggle.com/datasets/loveall/cervical-cancer-risk-classification>

<https://archive.ics.uci.edu/ml/datasets/Cervical+cancer+%28Risk+Factors%29#>