Cervical Cancer Risk Factors for Biopsy

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.

Front-End (Angie)

Create a site to display our plotly visualizations, links, etc.

Databricks (Estela)

Our model will utilize data retrieved from a big data source using Spark.

Model: Logistical Regression (Paola)

Clean, normalize, and standardize data prior to modeling

Initialize, train, and evaluate our model

Measure our model’s predictive power by measuring classification accuracy and R-squared

Presentation (Ramiro)

Source: Dataset was originally obtained from UCI Repository

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