


Cervical Cancer Risk Factor Analysis & Classification



Hunter Sapienza



Problem Statement

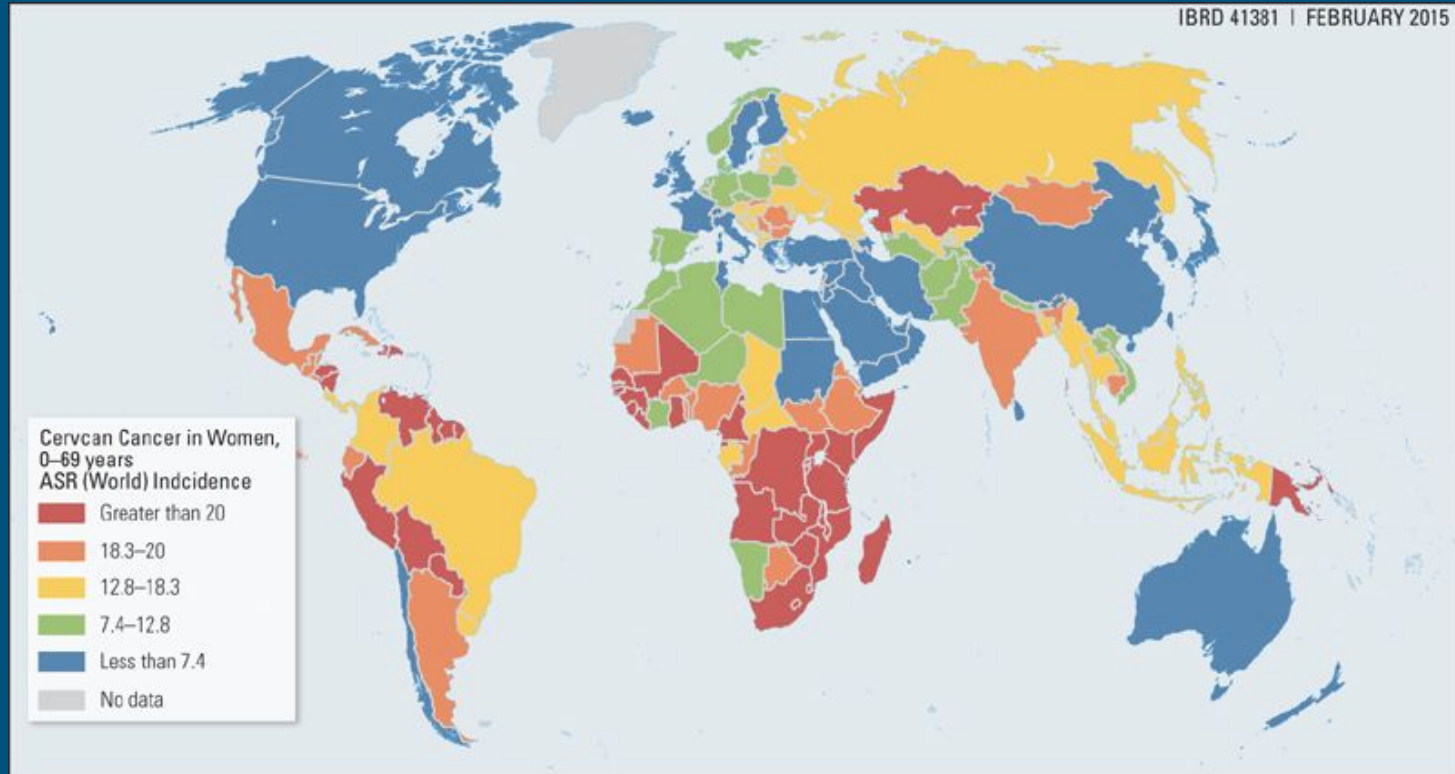
- What risk factors most significantly correlate with positive biopsies for cervical cancer?
- Which classification model performs most accurately while minimizing false negatives?

Business Value

How can we optimize patient care?

- Most accurate diagnoses prior to biopsy
- Minimize false negatives
- Implement preventative measures

Cervical Cancer Worldwide



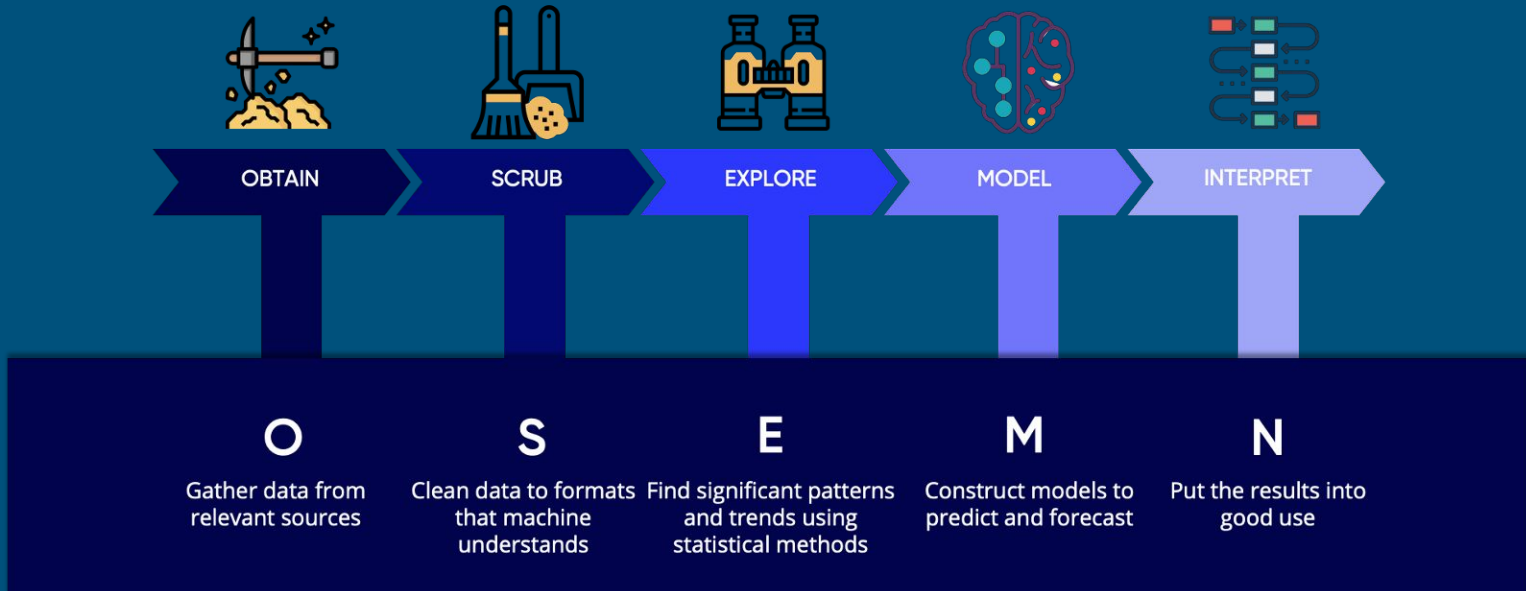
Cervical Cancer Risk Factor Data



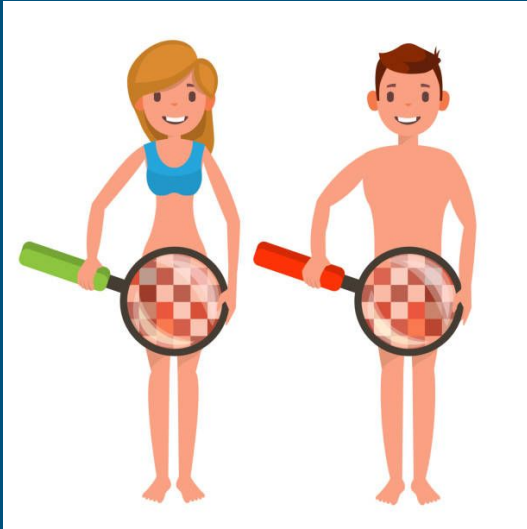
Hospital Universitario de Caracas

Data Science Framework

Data Science Process



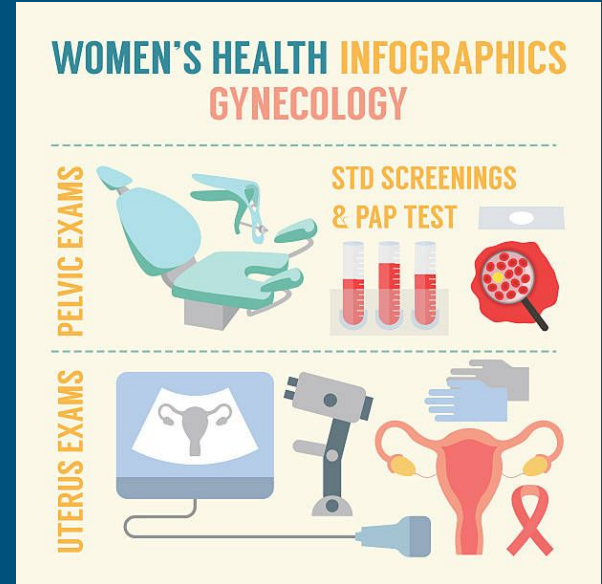
Top Predictors



Prior STDs

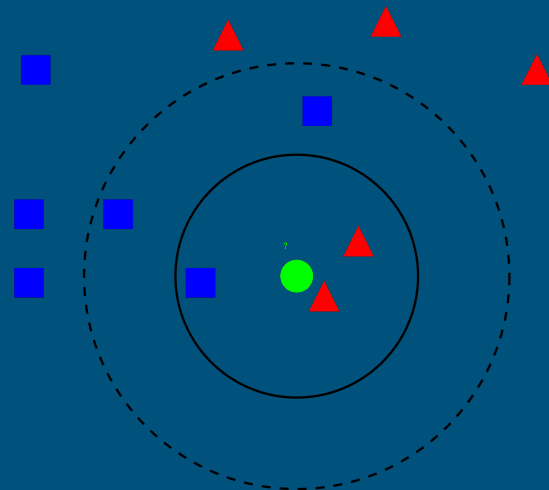
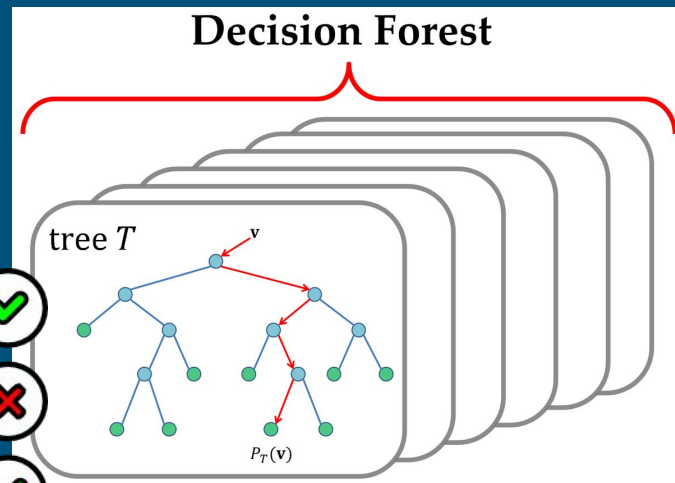
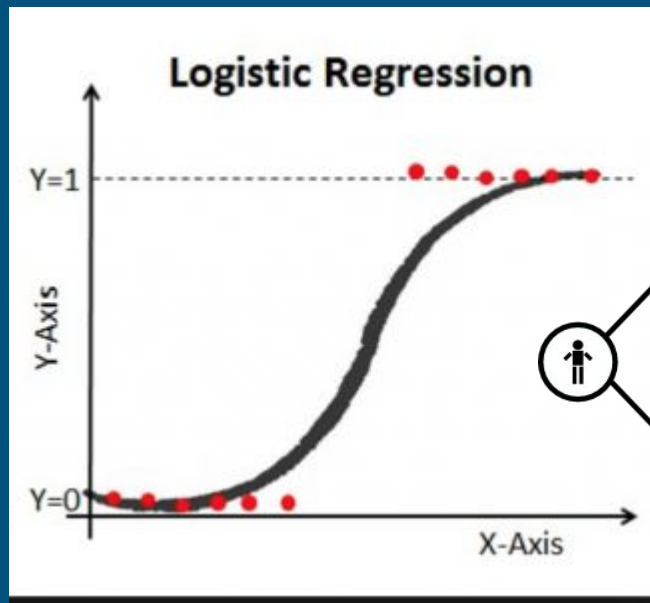


HIV, HPV

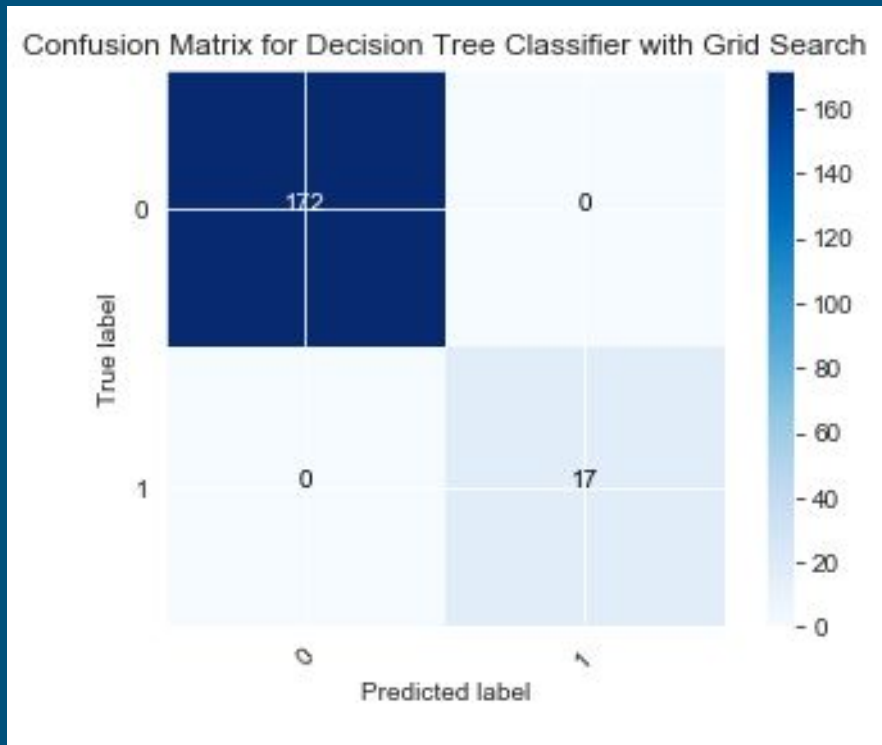


Other Tests

Modeling the Data



Modeling Findings



Decision Tree Classifier

Accuracy: 1.00

Recall: 1.00

Precision: 1.00

False Positives: 0

False Negatives: 0

Business Recommendations

- Increased HIV and HPV prevention measures
- Proactive STD testing and treatment
- Regular testing for susceptible women
- Thorough medical screening for risk factors



Future Work

1. Transferability to data from other countries
2. Multiclass classification model and analysis
3. Classifier model comparison for Hinselmann, Schiller, Citology, and Biopsy features
4. Classification models for Biopsy in absence of the other three diagnostic tests

