

Cervical Cancer Screening

Kirsty Hawke



Early Detection & Prevention

- Zambia is ranked 3rd with rates of 66/100,000¹
- One of the most preventable cancers²



Background

- Cervical cancer is a disease in which cancerous cells form in the tissues of the cervix
- Usually no obvious symptoms
- Detection: regular check-ups



Data

- Cervical cancer (Risk Factors)
Dataset (source)
- Collected in Venezuela in 2017
- 32 features from 858 patients
 - Demographic information
 - Habits
 - Medical information
- Issues
 - Many missing values (max 787 responses)
 - Small dataset
- Target
 - Biopsy result

Data Exploration

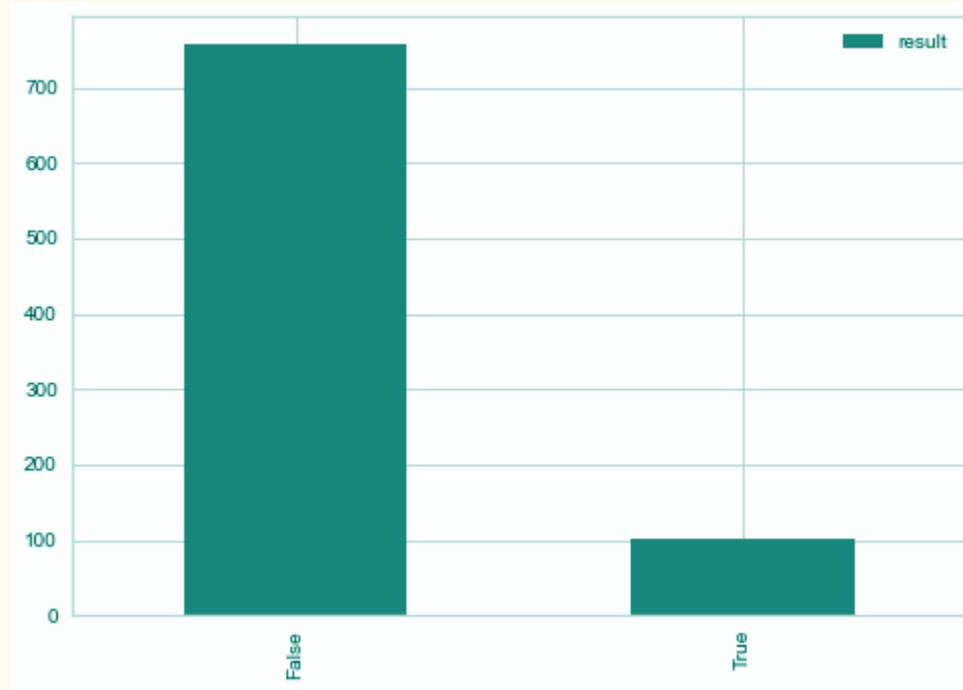


Fig 1: Class Distribution of Cervical Cancer Diagnosis

Cleaning Methods

- Resampling
 - Up-sampled positive class
- Imputation
 - K Nearest Neighbour – for missing values
- Scaling
 - RobustScaler - to reduce outlier effect



Feature Engineering - Feature Importance

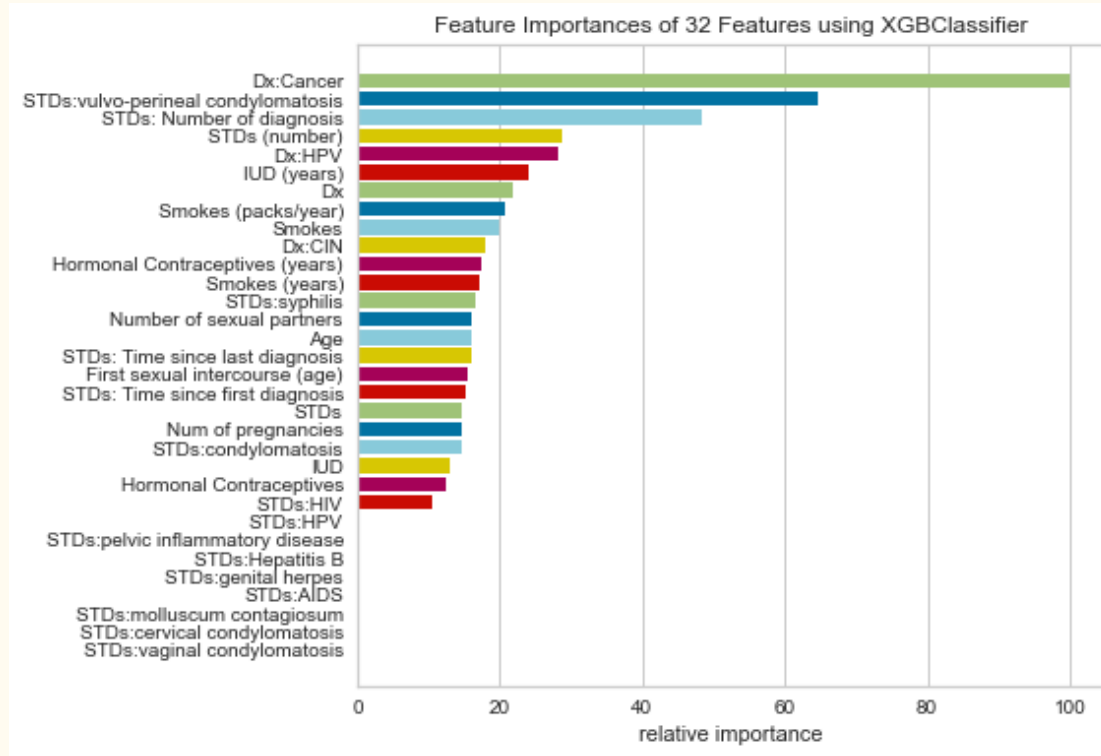
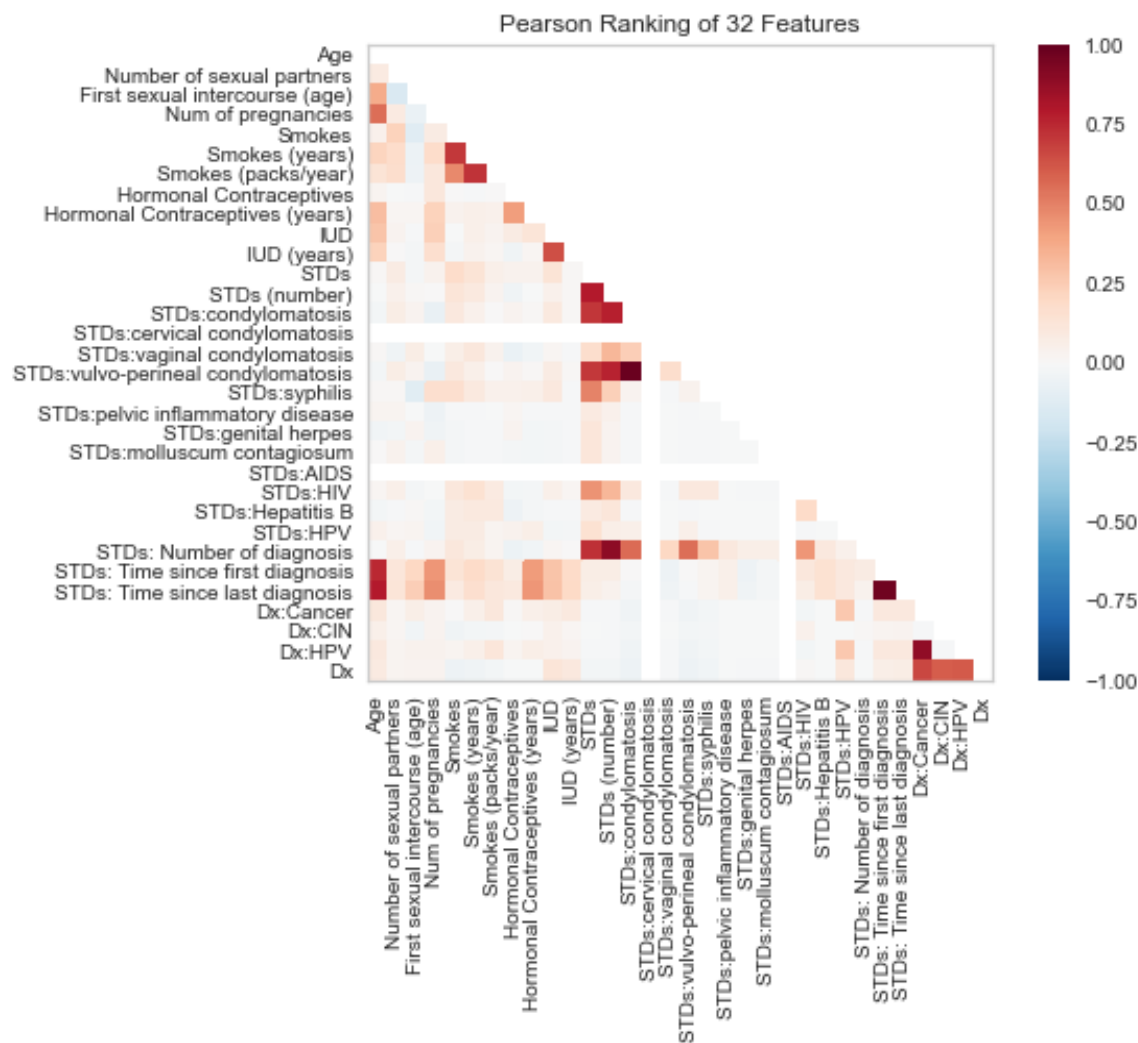


Fig 2: Feature Importance

Feature Engineering

Fig 3: Pearson Ranking of Features



Selected Features

Age

Number of pregnancies

Smokes (packs/year)

Number of sexual partners

First sexual intercourse (age)

Hormonal Contraceptives (years)

IUD (years)

STDs (number)

Model

- XGBoost Model
 - eXtreme Gradient Boosting
 - Speed & performance
- Baseline accuracy: 0.534
- Baseline F1 Score: 0.23
- F1 Score: 0.832
- Accuracy: 0.93

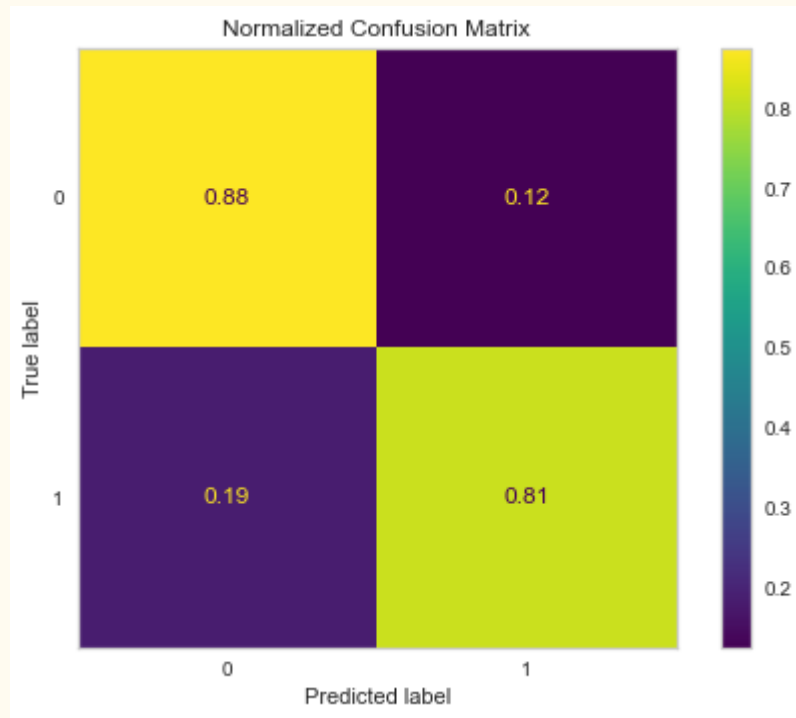
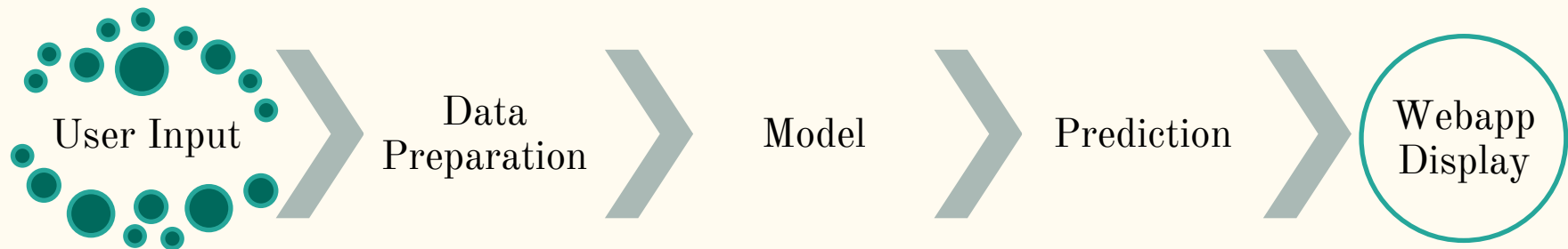


Fig 4: Confusion Matrix

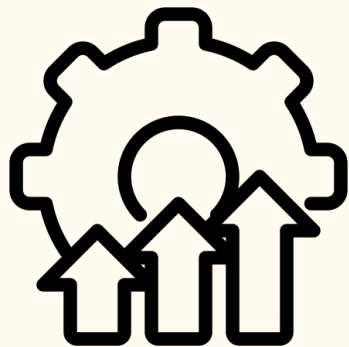
Workflow



Demo

Improvements

- Deploy webapp via Heroku
- Bigger dataset
 - Include whether or not the patient is vaccinated for HPV
 - To reduce overfitting
- Increased testing
- More appealing user interface
- Conduct pilot study to gather user feedback



Thank you!



<https://github.com/hawkek>

Tech Stack Used

- Python
- Streamlit
- Jupyter Notebook

Citations

1. <https://www.wcrf.org/dietandcancer/cervical-cancer-statistics/>
2. <https://www.kucancercenter.org/outreach/prevention/preventable-cancers>
3. <https://www.healthywomen.org/your-health/prevention--screenings/cervical-cancer-screening-save-life>
4. [Image](#)
5. Improvement by Kuku Wachyu Bias from NounProject.com

Overfit Model

