

PREDICTIVE ANALYTICS WITH MACHINE LEARNING

Predicting Cervical Cancer Risk

Data Source :

UCI Machine Learning Repository

Data Type :

Hospital Patient Data Set

Application :

Analance



PROOF OF CONCEPT – POWERED BY ANALANCE

At one point, **Cervical Cancer** was considered one of the most common causes of cancer death for American women. While mortality rate has dropped significantly due to an increase in screening, it has not changed much in the last decade.¹ To improve outcomes, timely intervention is critical.



LEVERAGING ML TO ENABLE EARLY DIAGNOSIS

Analance used a hospital's patient data set to predict the likelihood of a patient being diagnosed with cervical cancer—with summaries and findings easy to explore through dashboards and reports.

By leveraging Analance machine learning (ML), patients can be classified into risk categories based on their likelihood of having the disease and future outcomes can be predicted with accuracy as high as 94%. With built-in automations, alerts can be scheduled to notify healthcare professionals when patients are at risk of cervical cancer. This allows providers to implement proactive measures and improve outcomes.



EXPLORATORY DATA AND MODELING PROCESS

With visibility into a patient's risk of cervical cancer, providers can facilitate early detection, offer prompt treatment, or put measures in place to promote timely intervention.

A total of 858 patients were observed and 35 different predictor variables were considered such as age, number of sexual partners, age during first intercourse, number of pregnancies, presence of STDs, and more.

All variables available were studied to understand distributions. Data was cleaned by the means of handling outlying values, missing values, and looking for interrelationships between predictors before looking to see if any data had a significant relationship with the outcome. A Bivariate Analysis (Chi-Squared) was done for all predictoroutcome combinations, which helped in restricting the analysis to only those predictors that majorly influence cervical cancer risk.



DATA MODELING AND FINDINGS

A total of 50 different models were built but the Two-Class Adaptive Boosting model was chosen as the winning model based on the model accuracy. From the analysis performed, the highest risk of cervical cancer was found for:

- Patients who had their first sexual intercourse between ages 10 and 20 (96.4%)
- Patients who had at least 1 pregnancy (85.5%)



DATA ANALYSIS AND INSIGHTS

Patients who began to be sexually active at a younger age have a higher risk of cervical cancer. Researchers attribute this to cervix changes during puberty, which can make the area more vulnerable to damage. In fact, sexual activity generally increases one's risk of cervical cancer, since it potentially exposes patients to a human papillomavirus (HPV) infection (which is closely linked to the disease as well). Giving birth also increases risk. Research has suggested that the risk may be attributed to hormone changes during pregnancy, or the trauma to the cervix during birth.



NEXT STEPS

Providers can proactively request screening tests, recommend preventive care, and provide sex education and valuable medical advice to patients likely to develop cervical cancer.

SOURCES:

1. Cancer.org, "[Key Statistics for Cervical Cancer](#)"

ABOUT DUCEN

Ducen is a trusted technology solutions provider that aims to empower Fortune 1000 companies through quality solutions and services. We work with organizations across various verticals to drive business outcomes and enhance their customer experience. Through our enterprise analytics platform, we build and manage data-driven digital platforms including business intelligence and advanced analytics solutions. We also offer a comprehensive services portfolio covering data management and cyber security to help clients stay ahead of the technology curve.

For more information please visit www.analance.com



A Sryas Company