

INDIVIDUAL ASSIGNMENT 30%

Data Science Report: Understanding the Impact of Cervical Cancer

Executive Summary:

The Cervical Cancer dataset provides valuable insights into the factors influencing the prevalence and diagnosis of cervical cancer. By employing data science techniques, this analysis aims to explore key features, identify patterns, and provide recommendations for public health interventions. Please Provide Full Report, Presentation Slides and Recorded Presentation Video.

1. Introduction:

The objective of this analysis is to leverage data science methodologies to gain a comprehensive understanding of cervical cancer. The dataset contains information about various factors such as age, sexual history, smoking habits, contraceptive use, and the presence of sexually transmitted diseases.

2. Data Preprocessing (10 marks):

2.1 Handling Missing Values:

Identify and handle missing values in the dataset, ensuring data completeness for accurate analysis. (If Necessary)

2.2 Data Cleaning:

Remove duplicate entries and irrelevant columns to streamline the dataset. (If Necessary)

3. Exploratory Data Analysis (15 marks):

3.1 Descriptive Statistics:

Provide summary statistics for key variables, including age, number of sexual partners, and the prevalence of cervical cancer.

3.2 Temporal Analysis:

Visualize the trends and patterns in cervical cancer cases over time, if applicable.

4. Modeling and Prediction (50 marks):

4.1 Time Series Forecasting:

If applicable, use time series analysis to predict future trends in cervical cancer cases.

4.2 Machine Learning Model:

Build a machine learning model to predict the likelihood of cervical cancer based on relevant features(Hinselmann, Schiller, Citology, Biopsy) Pick total 3 Machine Learning Models by importing from sklearn.

4.3 Evaluation Metrics:

Assess the performance of the model using appropriate evaluation metrics, considering factors like precision, recall, and accuracy.

Apply all the models into every metrics, and according to the metrics, Visualize the plot for all models.

5. Interpretation of Results (15 marks):

5.1 Insights:

Summarize key insights obtained from the analysis, including factors strongly correlated with cervical cancer and potential risk factors.

6. Conclusion (10 marks):

In conclusion, this data science analysis provides valuable insights into the dynamics of cervical cancer. By leveraging advanced analytical techniques, we can better understand the factors influencing the prevalence of cervical cancer and make informed decisions for preventive measures and healthcare interventions.

Note: The Dataset Provided.