**MS in Applied Data Science - Portfolio**

**Name:** Sukhad Dnyanesh Joshi

* Course: IST 707 - Applied Machine Learning

Credibility Detection of Health Web Blogs Using Explainable AI: -

**Project Overview:**

This project aims to detect the credibility of online health blogs using supervised machine learning models. Techniques like Random Forest, XGBoost, and Explainable AI methods (SHAP values) were used to ensure that model predictions are transparent and understandable. The project addresses ethical AI challenges by making machine learning decisions interpretable for non-technical stakeholders. Through this approach, it highlights the importance of responsible AI practices when dealing with sensitive information like health data.

**Files in This Folder:**

|  |  |
| --- | --- |
| **File** | **Description** |
| Content\_analysis.ipynb, Extraction\_of\_weblinks.ipynb, customer\_ratings.ipynb, regression\_score.ipynb,  web\_scraper.py,  NibblerScraper.py, etc. | Jupyter Notebooks for model training and explanation |
| HealthBlog Credibility detection.csv | Dataset containing features of health blogs |
| SHAP Visualizations (separate images) | SHAP plots for model explainability |
| README.docx (IST 707\_Walkthrough) | Project documentation (this file) |

**GitHub Repository:**

<https://github.com/SukhadJoshi/MS-ADS-Portfolio_Sukhad-Dnyanesh-Joshi>

**Software Requirements:**

* Python 3.10 or later
* Jupyter Notebook
* Libraries: pandas, scikit-learn, xgboost, shap, matplotlib, seaborn

**How to Review:**

1. Open all the files in the suitable IDE such as VS Code or using Jupyter Notebook.
2. Run all the files and explore SHAP visualizations explaining model predictions and scores for the credibility of blogs.
3. Then we can validate and check the score of these blogs to determine how machine learning can be made more transparent for stakeholders.