

1. Title:

Anemia Detection with Machine Learning

2. Abstract:

- Anemia is a common condition caused by a lack of healthy red blood cells or hemoglobin, resulting in symptoms like fatigue and weakness.
- Early detection is crucial for effective treatment and better health outcomes.
- This project aims to develop a machine learning model for anemia detection using a dataset from Kaggle provided by Biswa Ranjan Rao.
- MLOps practices are incorporated to ensure efficient deployment and management of the model.

3. Overview:

- The dataset includes 1,421 samples with features such as gender, hemoglobin level, MCH, MCHC, MCV, and a binary outcome (anemic or not).
- Focus is on building a simple, yet effective predictive model.
- Key aspects include data preprocessing, model training, evaluation, and applying MLOps for deployment.

4. What's Inside:

- Data Analysis: Exploration of data to understand distributions and patterns.
- Feature Selection: Identifying the most important features that influence anemia.
- Class Imbalance Handling: Techniques like oversampling and undersampling to balance the data.
- Model Training: Algorithms like Decision Tree, Random Forest, Logistic Regression, K-Nearest Neighbors, SVM, and Naive Bayes.
- Evaluation: Using metrics such as accuracy, precision, recall, F1 score, and AUC.
- MLOps Practices: Implementing MLOps for model deployment, monitoring, and scalability.

5. Programming Language Used:

- **Python:** Primary language for model building and deployment.
 - Utilizes libraries such as Scikit-learn, Pandas, Matplotlib, and Seaborn.

6. Requirements:

➤ Software:

- Python 3.x
- Jupyter Notebook or any Python IDE

➤ Libraries:

- Scikit-learn for machine learning algorithms.
- Pandas for data manipulation.
- Matplotlib and Seaborn for data visualization.

➤ MLOps Tools:

Tools like Docker, Git, and CI/CD pipelines (GitHub Actions).

7. Database Used:

➤ Dataset:

Consists of attributes such as gender, hemoglobin, MCH, MCHC, MCV, and anemia status.

8. Algorithms Used and Other Usages:

Algorithms:

- Decision Tree
- Random Forest
- Logistic Regression
- K-Nearest Neighbors (KNN)
- Support Vector Machine (SVM)
- Gaussian Naive Bayes

Other Usages:

- Data Preprocessing: Cleaning and preparing data for model training.
- Hyperparameter Tuning: Grid search and cross-validation for optimizing model performance.
- Model Deployment: Using MLOps tools for deploying and managing the model in production environments.