Technology : LightGBM is used which is a gradient boosting framework that uses tree-based learning algorithms.

Implementation :

1) Irrelevant columns (respondent\_id, employment\_industry, employment\_occupation, health\_insurance) are dropped from the feature sets.

2) Numerical and categorical features are identified from the training feature set.

3) Numerical Features: Imputed using IterativeImputer and scaled using StandardScaler.

4) Categorical Features: Imputed using SimpleImputer and one-hot encoded using OneHotEncoder.

5) Both transformers are combined into a ColumnTransformer to preprocess the features appropriately.

6) A StratifiedKFold splitter is used to maintain the distribution of the target variable in each fold.

7) A function evaluate\_model is defined to train a LGBMClassifier model using the preprocessed data and compute the ROC AUC score for each fold.

8) The function evaluate\_model is executed for xyz\_vaccine, Seasonal\_vaccine and the mean ROC AUC score is printed.

9) An overall ROC AUC score is computed as the average of the two vaccine scores.

10) A DataFrame containing the respondent\_id and the predicted probabilities for both vaccines is created and saved to a CSV file named results.csv.

Result : ROC AUC=0.84822 with narrow confidence intervals and low standard deviations

(-+0.01). This suggests

stability in the models’ ability to discriminate between positive and negative samples for both target variables.