**Predictive Modeling of Vaccination Status: A Comparative Study of Machine Learning Algorithms**

**1. Introduction**

Vaccination against infectious diseases is one of the most effective public health interventions. In this project, we aim to classify whether an individual is vaccinated or not based on various features. This classification task is crucial for understanding vaccination coverage and identifying populations that may need targeted interventions to improve vaccination rates.

**2. Dataset**

The dataset used in this project contains information about individuals, including demographic characteristics, medical history, and vaccination status. Each record represents a person, and the target variable indicates whether they are vaccinated or not.

**3. Methodology**

We employed several machine learning algorithms for classification:

- Logistic Regression

- Decision Tree Classifier (DTC)

- Bagging Classifier

- Adaboost Classifier

- Gradient Boosting Classifier

- Random Forest Classifier

- Support Vector Classifier (SVC)

For each algorithm, we compared the performance of both tuned and untuned models. Tuning involved optimizing hyperparameters using techniques such as grid search or random search.

**4. Feature Engineering**

Before training the models, we performed feature engineering, including handling missing values, encoding categorical variables, and scaling numerical features. This step ensured that the data was suitable for modeling.

**5. Model Evaluation**

We evaluated the models using the following metrics:

- Accuracy: Overall correctness of the predictions

- Precision: Proportion of true positive predictions among all positive predictions

- Recall: Proportion of true positive predictions among all actual positives

- F1-Score: Harmonic mean of precision and recall

**6. Conclusion**

In conclusion, we successfully developed models to classify individuals' vaccination status using various machine learning algorithms. Among the models compared, Gradient Boosting and Random Forest classifiers demonstrated the highest performance in terms of accuracy and F1-Score. Tuning the hyperparameters improved the performance of some models, indicating the importance of hyperparameter optimization.

**7. Future Work**

In future iterations of this project, additional features could be explored to further enhance model performance. Additionally, ensemble methods such as stacking could be investigated to combine the strengths of multiple models for improved predictions.

Overall, this project provides valuable insights into predicting vaccination status and underscores the significance of machine learning in public health research and interventions.