**ABSTRACT FOR PARTICIPATION IN KAPS/ INTER-UNIVERSITY 2023 HACKATHON**

**Topic: Data Analytics and Machine Learning**

**Flu Shot Learning: Predict H1N1 (Swine Flu) and Seasonal (Regular Flu) Vaccines**

**Team:**

1. **Njoki Ruth Olima**
2. **Mokua James Edgar**

The goal of this study will be to predict the likelihood of individuals receiving both H1N1 (Swine Flu) and seasonal flu vaccines (Regular Cold) based on a National 2009 H1N1 Flu Survey conducted in the USA. This prediction involves two binary target variables: 'h1n1\_vaccine' and 'seasonal\_vaccine,' where 0 signifies no vaccination, and 1 indicates vaccination.

Performance evaluation of the model will be based on the area under the receiver operating characteristic curve (ROC AUC) for both 'h1n1\_vaccine' and 'seasonal\_vaccine.'. The overall prediction score of our model will be determined by the mean of these two ROC AUC scores.

A higher score reflects superior predictive performance, with Python's **sklearn.metrics.roc\_auc\_score** being the most suitable tool to calculate these scores, using the 'macro' parameter for averaging in this multilabel setup. This study aims to provide valuable insights into factors influencing vaccination decisions for H1N1 and seasonal flu, enabling better public health planning and interventions. \*The COVID-19 pandemic forms the basis for this study. Unfortunately, getting data on COVID-19 is difficult as it is still a sensitive topic

The dataset contains 36 columns, with 'respondent\_id' as the unique identifier and 35 features. These features encompass various aspects such as the level of concern and knowledge about H1N1, behavioral factors (e.g., medication use and handwashing), doctor recommendations, health conditions, demographics (age, education, race, sex), and socio-economic factors (income, marital status, housing situation, employment status). Notably, the prediction task is formulated as a multilabel problem.