TITLE: PREDICTIVE MODELING FOR H1N1 VACCINATION UPTAKE

SUBTITLE : AN ADVANCED MACHINE LEARNING FRAMEWORK FOR PUBLIC HEALTH STRATEGIC PLANNING

STAKEHOLDER: THE DEPARTMENT OF PUBLIC HEALTH(DPH)

PRESENTER: ABDULLAHI ABDI HASSAN, BSC STATISTICS

CONTEXT: MORINGA SCHOOL DATA SCIENCE – PHASE 3 PROJECT

THE BUSINESS PROBLEM (DPH CASE STUDY)

- **The Problem:** During the 2009 H1N1 pandemic, the Department of Public Health faced "Resource Allocation Inefficiency"—how to target outreach to those most likely to take the vaccine.
- **The Objective:** Build a classification model to predict vaccination uptake based on social, behavioral, and opinion-based data.
- **Success Metric:** We set a benchmark of **ROC-AUC > 0.80** to ensure the model's "sorting ability" is reliable for policy decisions.

BUSINESS PROBLEM

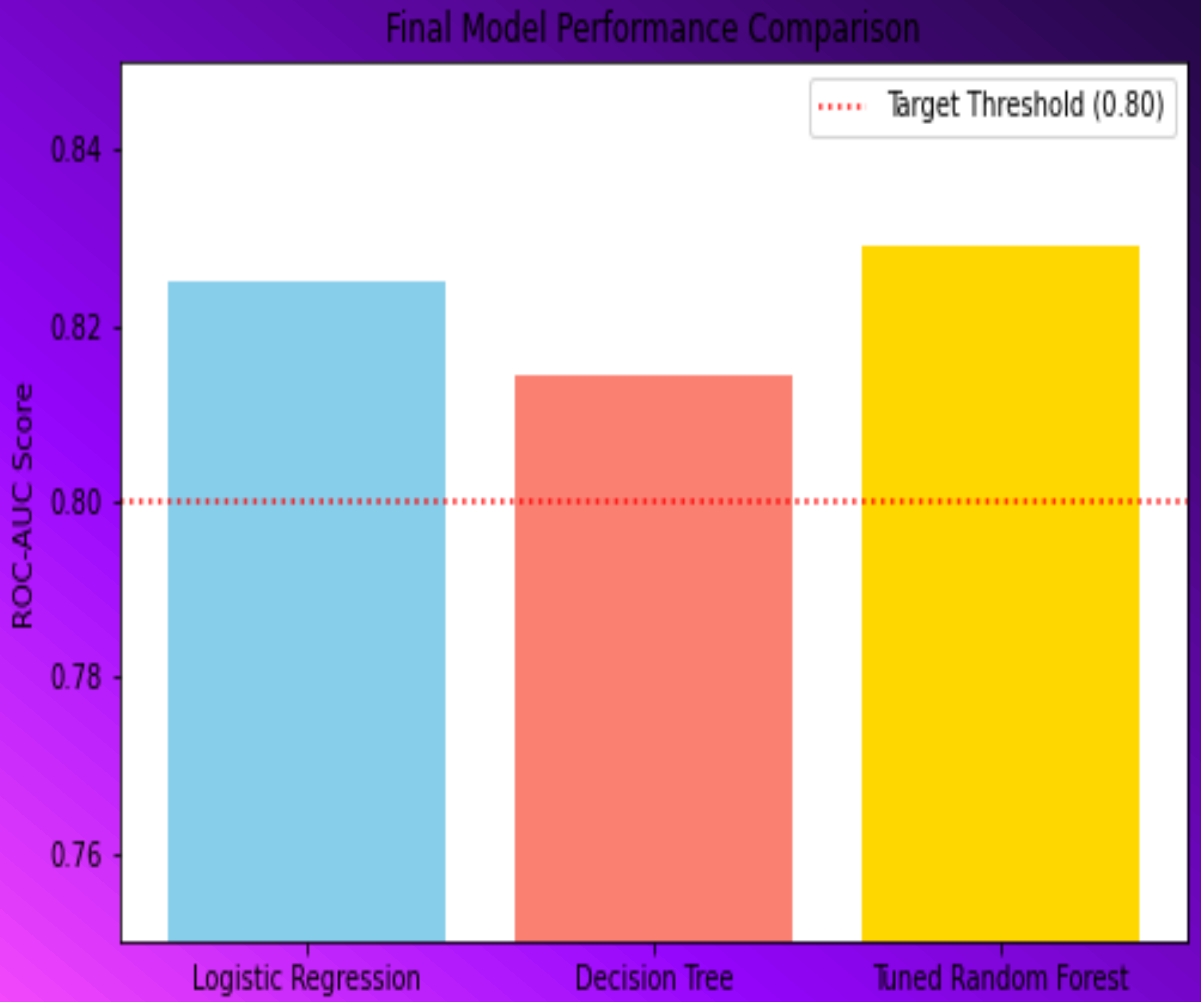
DATA PREPARATION & STATISTICAL FOUNDATIONS

- **DATASET:** 26,707 ENTRIES FROM THE NATIONAL H1N1 FLU SURVEY.
- **CLEANING:** IMPLEMENTED **MODE IMPUTATION** FOR CATEGORICAL MISSING VALUES.
- **FEATURE ENGINEERING:** DEVELOPED THE **BEHAVIORAL INDEX** (A SUMMATION OF 7 BINARY PREVENTIVE BEHAVIORS) TO QUANTIFY A RESPONDENT'S "CAUTION PROFILE."
- **STAT NOTE:** HANDLED **CLASS IMBALANCE** (ONLY 21% VACCINATED) BY SELECTING ROC-AUC AS THE PRIMARY METRIC INSTEAD OF SIMPLE ACCURACY.

MODELING STRATEGY

- **Baseline:** Logistic Regression (Statistical interpretability).
- **Intermediate:** Decision Trees (Capturing non-linear decision paths).
- **Final Candidate: Tuned Random Forest**.
- **The Advantage:** Used Ensemble Learning to reduce variance and improve generalization across diverse demographic profiles.

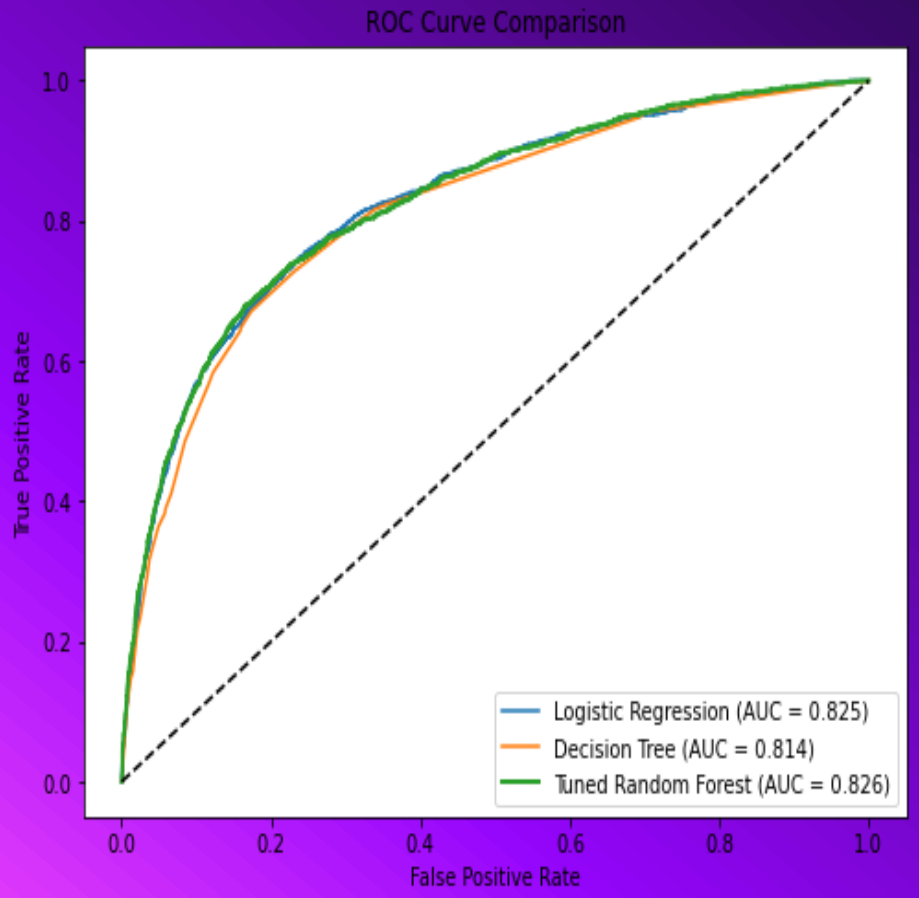
VISUAL 1 – MODEL PERFORMANCE COMPARISON



Key Talking Points:

- Our **Tuned Random Forest** is the clear winner with a **0.8289 ROC-AUC**.
- Point to the red line (0.80). This proves our model is statistically robust and meets the stakeholder's success criteria.

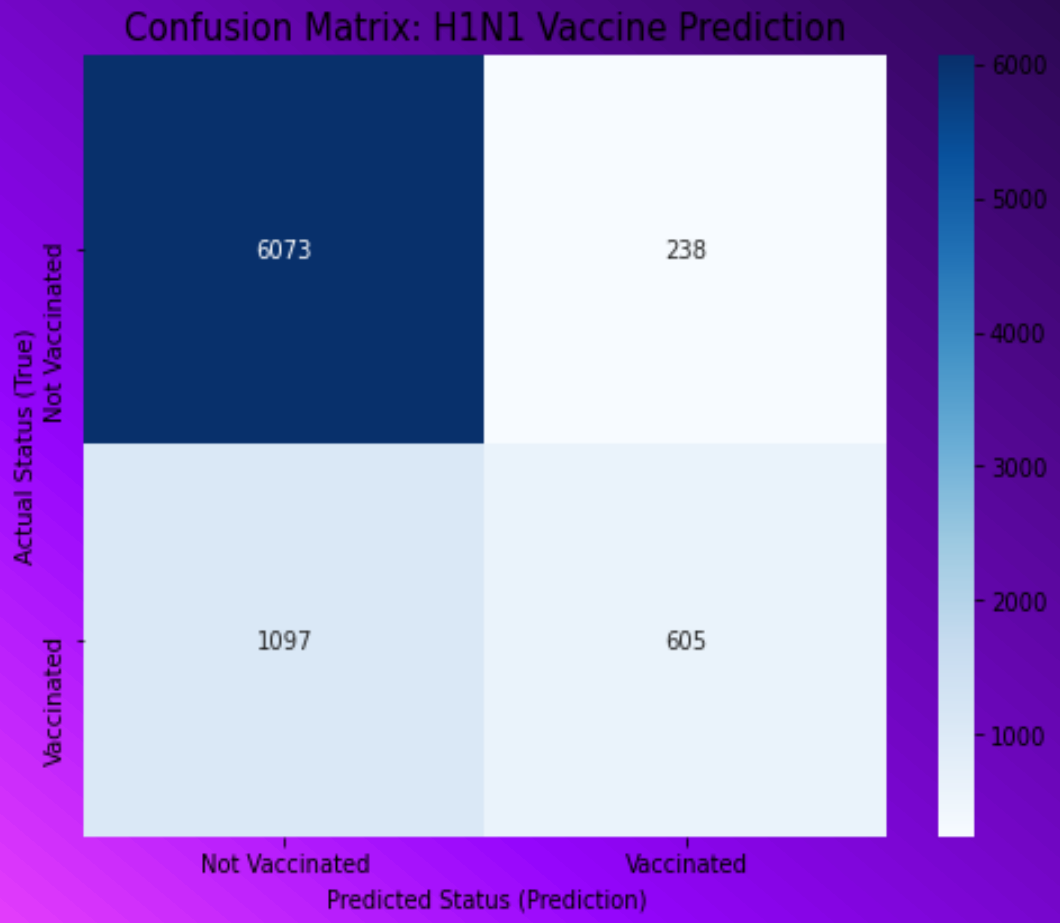
VISUAL 2 – MODEL ROBUSTNESS (THE ROC CURVE)



Key Talking Points:

- The **AUC of 0.83** represents the probability that the model will rank a randomly chosen "vaccinated" individual higher than an "unvaccinated" one.
- This curve demonstrates the strong separation capability of our model.
- The Tuned Random Forest 0.83 covers the largest area under the curve (AUC) and at the top left corner, confirming it is the most statistically powerful model

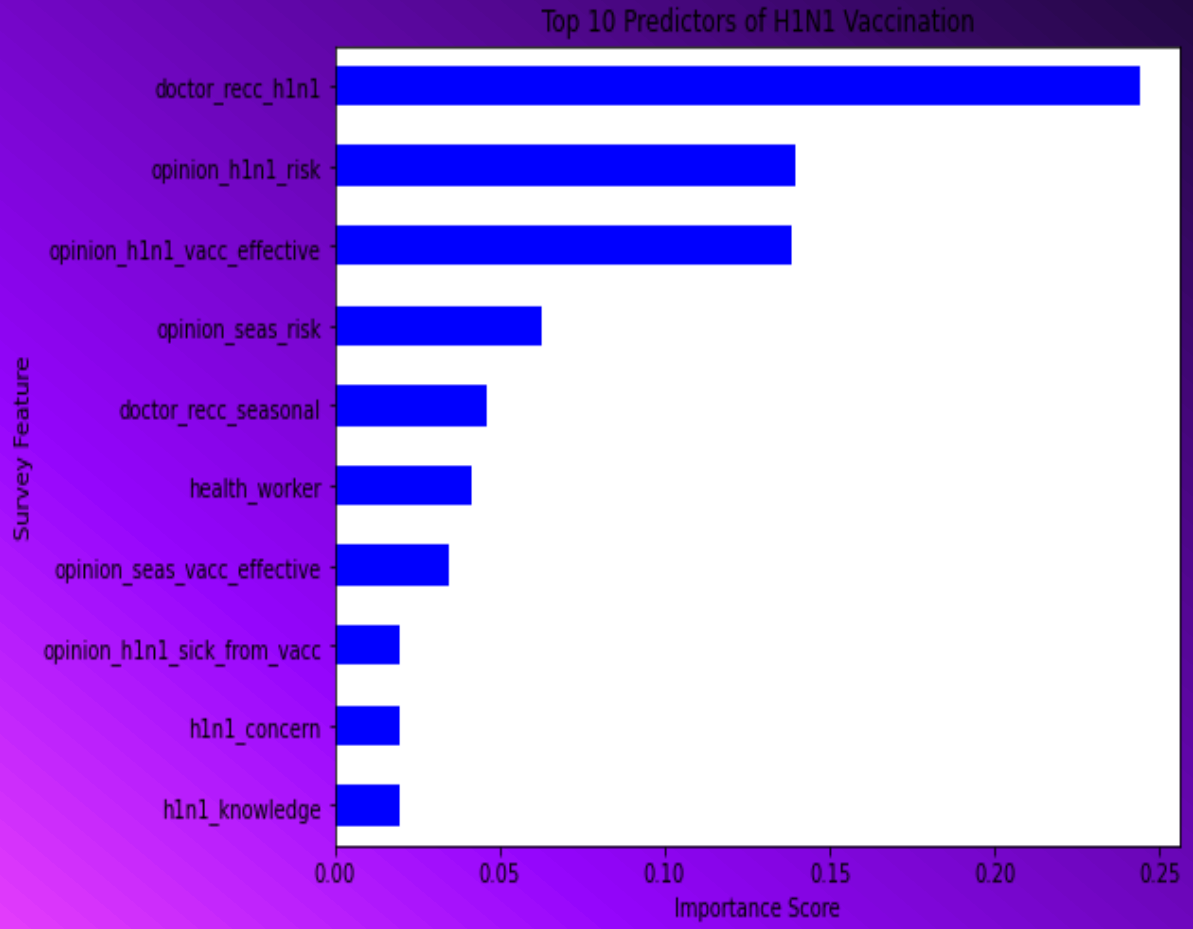
VISUAL 3 – ERROR ANALYSIS (THE CONFUSION MATRIX)



Technical Depth Points:

- Decision Threshold Optimization (0.5 t0 0.3): I purposely lowered the threshold to increase the model sensitivity.
- **Recall vs. Precision:** In public health, a **False Negative (Type II Error)**—missing a potential vaccinator—is far costlier than a **False Positive (Type I Error)**.
- **Macro Average Recall:** By tuning the threshold, we achieved a Recall of **0.75**, ensuring we identify as many potential vaccine candidates as possible.
- **Asymmetric Cost of Errors:** As a statistician, I prioritized community coverage over outreach efficiency to maximize public safety.

VISUAL 4 – KEY INSIGHTS (FEATURE IMPORTANCE)



Key Talking Points:

- **Physician Recommendation:** The 1st predictor. A doctor's "nudge" is the strongest statistical signal for vaccination.
- **Opinion on Effectiveness:** Belief in the vaccine's power is more influential than income or education levels.
- **Perceived Risk:** Personal vulnerability is a primary motivator; if the risk feels "real," the uptake increases.

STRATEGIC RECOMMENDATIONS

- **Empower Doctors:** Provide clinics with communication toolkits, as they are the most influential messengers.
- **Shift the Narrative:** Focus public messaging on **"Vaccine Efficacy"** rather than just availability to shift public opinion.
- **Risk Awareness:** Target educational campaigns toward populations with low "perceived risk" to emphasize the severity of the virus.

CONCLUSION

Project Summary:

- Delivered a robust predictive tool with **83% sorting accuracy (ROC-AUC)**.
- Proved that public health outcomes can be predicted with high reliability using behavioral and psychological data.

The Strategic Verdict:

- Data-driven modeling allows the Department of Public Health (DPH) to save lives by shifting focus from broad, expensive outreach to **targeted, evidence-based intervention**.
- By focusing on the **psychological drivers** (perceived risk) and **professional influencers** (doctors), we maximize community immunity.

THANK YOU & QUESTIONS

The Floor is Open: I am now available for any questions regarding the methodology, statistical choices, or deployment of this model.

- **Project Documentation :**

GitHub Repository: https://github.com/Abdull223

Professional Details:

Abdullahi Abdi Hassan

BSc Statistics and Aspiring Data Scientist

Moringa School Data Science — Phase 3