Neuro Shot:
A Flu Shot Learning Project:
Predict H1N1 and Seasonal
Flu Vaccines

By:

Sergii Makhkin

Diana Paola Americano Guerrero

Problem

- Data comes from the National 2009 H1N1 Flu Survey in the United States, so results should be considered within the context back then.
- However the importance of vaccination remains, especially in this post-pandemic world with anti-vaccination movements.
- Through this analysis we seek to know what factors (demographical, geographical, behavioral) influence vaccine uptake and can we use them to predict real world data.

Exploratory Data Analysis – 1

◆ **Demographics**

- Age is fairly balanced but slightly skewed towards older groups (65+).
- Sex: More females (59%) than males (41%).
- Race: Majority White (79%), with smaller representation of other groups.
- Income: Most respondents above poverty threshold.
- Employment: Majority employed (54%), ~41% not in labor force, small unemployed group.

◆ **Potential Challenges**

1. **Class imbalance** for H1N1 vaccine → may require resampling, class weights, or careful metric choice.
2. **High missingness** (~50%) in employment and insurance data → decisions needed: drop, impute, or encode "unknown".
3. **Mixed feature type**s (ordinal, categorical, binary, continuous) → preprocessing pipeline must handle appropriately.
4. **High-cardinality categories** (employment_industry, employment_occupation, hhs_geo_region) → need encoding strategies (target encoding, frequency encoding, etc.).
5. **Correlation among opinion variables** → risk of multicollinearity, might need dimensionality reduction or feature grouping.
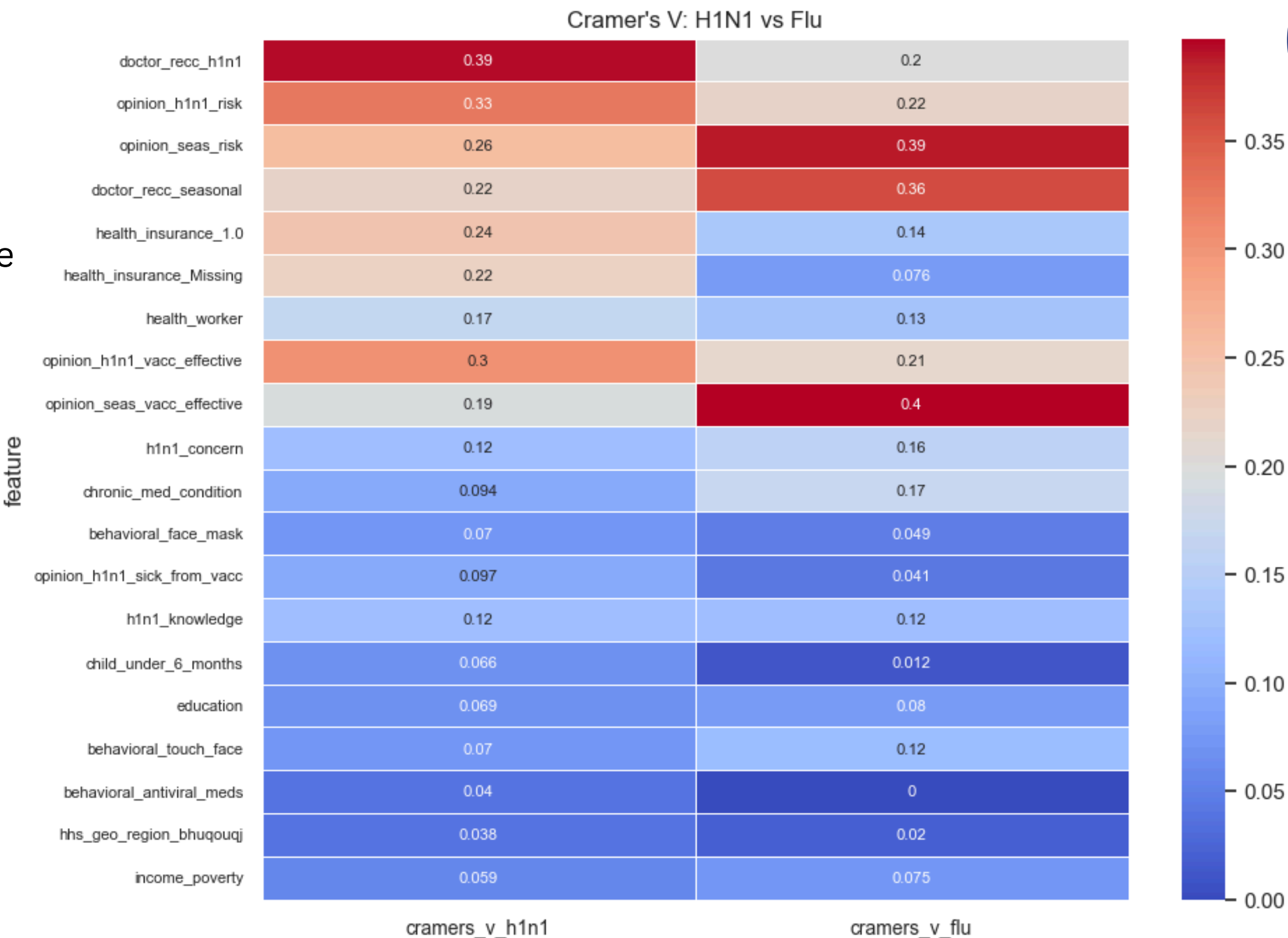
ReDI School of
Digital Integration

Exploratory Data Analysis – 2

◆ **Feature Analysis**

- Doctor recommendation is a shared important feature for both vaccines.
- Perceived personal risk and belief in vaccine effectiveness are also highly associated, health insurance also plays a moderate association.
- Demographics and geography have low association compared to opinion and behavior.
- Perceptions about one vaccine might influence uptake of the other.

→ Next step for feature selection is a model-based feature selection utilizing more complex models (ex. gradient boosting).

Cramer's V: H1N1 vs Flu

feature	cramers_v_h1n1	cramers_v_flu
doctor_recc_h1n1	0.39	0.2
opinion_h1n1_risk	0.33	0.22
opinion_seas_risk	0.26	0.39
doctor_recc_seasonal	0.22	0.36
health_insurance_1.0	0.24	0.14
health_insurance_Missing	0.22	0.076
health_worker	0.17	0.13
opinion_h1n1_vacc_effective	0.3	0.21
opinion_seas_vacc_effective	0.19	0.4
h1n1_concern	0.12	0.16
chronic_med_condition	0.094	0.17
behavioral_face_mask	0.07	0.049
opinion_h1n1_sick_from_vacc	0.097	0.041
h1n1_knowledge	0.12	0.12
child_under_6_months	0.066	0.012
education	0.069	0.08
behavioral_touch_face	0.07	0.12
behavioral_antiviral_meds	0.04	0
hhs_geo_region_bhuqouqj	0.038	0.02
income_poverty	0.059	0.075

ReDI School of Digital Integration

ReDI School of Digital Integration

Thank You.

For Your Attention

https://github.com/americano-diana/neuro_flushot/