Understanding emotional driving factors of vaccination hesitancy and refusal in Bangladesh using Machine Learning approaches

Supervised by

Adnan Ferdous Ashrafi,
(Senior Lecturer)
(Department Of Computer Science & Engineering)
Stamford University Bangladesh

Presented by

Abdullah Al Mamun
ID: CSE 070-08109
Batch: 70-A

Mahfuzur Rahman
ID: CSE 070-08111
Batch: 70-A

Content of this slide

01
INTRODUCTION

Introduction

- What is vaccine?
- Why do we need vaccine?
- Why should we got vaccinated?
- What is Emotional drivers?
- Relation between vaccination and emotional driver
- Motivation of this domain.
- Why do we choose this domain?
- Research outcome
- Types of tools and methods
  - Google Colab
  - Pandas
  - Numpy
  - Seaborn
  - Matplotlib

02

LITERATURE REVIEW

LITERATURE REVIEW

Walsh et. al [1]

Factors Influencing COVID-19 Vaccine Hesitancy and Acceptance

- Past vaccination behavior, risk perception, social influences and civic responsibility
- Government and healthcare providers

Rancher et. al [2]

Understanding COVID-19 Vaccine Hesitancy in the United States

- Race, ethnicity, socioeconomic status
- Open communication and equitable access to vaccination

Raut et. al [3]

Understanding COVID-19 Vaccine Acceptance and Hesitancy

- Misinformation, conspiracy theories and political polarization
- Vaccin literacy, trust in vaccin and government

LITERATURE REVIEW

Gupta et.al [4]

A Comprehensive Analysis of COVID-19 Dynamics: Insights from Statistical and Machine Learning Approaches

- Demographic, medical and vaccination
- Find the factors influence most

Saifuzzaman et. al [5]

Navigating the Impact of COVID-19: A Multifaceted Literature Review

- Brought unusual challenges
- Educational disruption, economic challenges,
- Mental health, reshaping the social norms

03

Methodology

Methodology

➔ **Dataset description**
- **Actual dataset [6]**
  - 500 rows(Men 250, women 250) &195 columns
  - Divided into sub-parts:

1. Descriptive & Demographic fields		
Domain of this subpart	**Description**	**Name of columns**
Descriptive fields	8 columns	(Id, survey_Start date ,survey_End date ,Duration in second, Info consent, redirect, finished, status_R)
Demographic fields	9 columns	(age, sex, district, religion, marital_status, education, education_other,conib, household)

Methodology

2. Individual Differences		
Domain of this subpart	**Description**	**Name of average columns**
The Big Five personality traits	comprises 30 items/columns,forming 5	(BFI_agree, BFI_cons, BFI_negam, BFI_extra, BFI_open)
Horizontal-Vertical,Individualism-Collectivism (HVIC)	comprises 14 items, forming four subscales	(HVIC_HC_avg, HVIC_VC_avg, HVIC_HI_avg, HVIC_VI_avg)
Future Consequences (CFC) scale	comprises 14 items/columns,forming 2 columns	(CFC_F_avg, CFC_I_avg)
General Prosocial tendencies	comprises 23 items/columns,forming single column	(prosoc_ALL_avg)

Methodology

3. Emotional responses & Threats		
Domain of this subpart	**Description**	**Name of average columns**
Feelings of helplessness related to the vaccination, the COVID-19 pandemic, and climate	Comprises 12 items/columns,forming 3 columns	(helplessness_CLIM_avg, helplessness_COV_avg, helplessness_VAC_avg)
Feelings of threat related to vaccination against the COVID-19 disease, the disease itself, and climate change	Comprises 9 items, forming 3 subscales	(threat_CLIM_avg, threat_COV_avg, threat_VAC_avg)

Methodology

4. Prosocial Intentions and Behaviour		
Domain of this subpart	**Description**	**Name of average columns**
Prosocial behavior	Comprises 21 items/columns,forming 3	(procovid_avg, proenviro_avg, provacc_avg)
Prosocial Altruistic	Comprises 5 items, forming single average column	(prosoc_ALT_avg)
Attitudes towards vaccination	comprises 10 items/columns,forming 5	(prosoc_EMO_avg,prosoc_PUB_avg,prosoc_DIR_avg,prosoc_COM_avg, prosoc_ANO_avg)
Anti vaccination average	comprises 4 items/columns,forming single column	(antivacc_avg)

Methodology

- **Validation dataset**
  - 48 response and 30 columns
  - Collected by google form
- **Data Preprocessing**
  - **Actual dataset**
    - Get 30 columns from 195 columns and made a new data frame for further operation.
    - Drop index column.
    - Target variable binning and converted data type as integer.
    - Converted categorical columns into integer value using lambda expression.
    - Applied min-max normalization for normalizing columns (Not on target column).
    - Applied feature selection technique and selected 20 columns for final operation.

Methodology

➔ **Data Preprocessing**

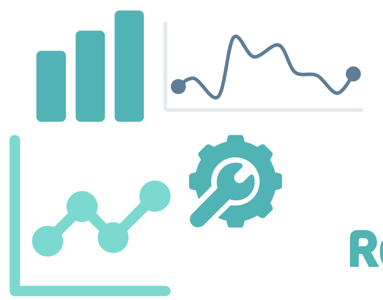
- **Validation dataset**
  - Drop 4 irrelative columns(Timestamp, name, profession,info_consent)
  - Use lambda function, mapping and astype function to convert categorical to numerical value
  - Applied min-max normalization technique to normalize all value
  - Visualized all columns (Boxplot, Histogram, scatterplot)
  - Then applied machine learning models.

04

Experimental Setup

Experimental setup

Techniques used for preprocessing	Models, Techniques & Algorithms used for result	Evaluation matrix
◆ Min-Max Normalization ◆ Round function ◆ Feature engineering	◆ Linear SVM ◆ Radial SVM ◆ Logistic Regression ◆ K-Neighbour Classifier ◆ Decision Tree classifier ◆ Gradient Boosting Classifier ◆ Random Forest Classifier ◆ Naive Bayes	◆ Accuracy ◆ Precision ◆ Recall ◆ F1-Score ◆ RSME



05

Result & Analysis

Result & Analysis

➔ K-fold cross validation

Model name	For 5-fold	For 10-fold	For 15-fold
Gradient Boosting Classifier	**0.98**	**0.98**	**0.97**
Decision Tree Classifier	0.94	0.96	0.96
Random Forest Classifier	0.93	0.92	0.93
Linear Svm	0.92	0.91	0.92
Logistic Regression	0.91	0.92	0.91
Gaussian Naive Bayes	0.90	0.89	0.89
Radial Svm	0.89	0.90	0.89
KNeighbors Classifier	0.78	0.77	0.77

Table 5.1: Performance of different classifiers using K-fold cross validation with 30 features

Model name	For 5-fold	For 10-fold	For 15-fold
Gradient Boosting Classifier	**0.98**	**0.98**	**0.98**
Decision Tree Classifier	0.95	0.95	0.95
Random Forest Classifier	0.93	0.93	0.93
Linear Svm	0.93	0.93	0.93
Logistic Regression	0.91	0.91	0.91
Gaussian Naive Bayes	0.94	0.93	0.93
Radial Svm	0.90	0.91	0.90
KNeighbors Classifier	0.77	0.79	0.77

Table 5.2: Performance of different classifiers using K-fold cross validation with 21 features

Result & Analysis

➔ Split ratio and accuracy

Model name	Split ratio		
	85:15	80:20	70:30
Gradient Boosting Classifier	**0.97**	**0.97**	**0.98**
Linear Svm	0.92	0.91	0.90
Decision Tree Classifier	0.90	0.95	0.93
Random Forest Classifier	0.89	0.89	0.91
Logistic Regression	0.85	0.90	0.88
Radial Svm	0.84	0.88	0.85
Gaussian Naive Bayes	0.81	0.86	0.85
KNeighbors Classifier	0.80	0.84	0.82

Table 5.3: Performance of different classifiers with different data split ratios for 30 features

Model name	Split ratio		
	85:15	80:20	70:30
Gradient Boosting Classifier	**0.98**	**0.97**	**0.97**
Linear Svm	0.91	0.92	0.90
Decision Tree Classifier	0.90	0.95	0.93
Random Forest Classifier	0.90	0.93	0.93
Logistic Regression	0.88	0.91	0.87
Radial Svm	0.85	0.92	0.87
Gaussian Naive Bayes	0.85	0.89	0.88
KNeighbors Classifier	0.82	0.86	0.80

Table 5.4: Performance of different classifiers with different data split ratios for 21 features

Result & Analysis

➔ Individual performance matrix and hyper parameter tuning

Model name	F1_score		rmse	accuracy_score	
	Train_data	Test_data		Train_data	Test_data
GradientBoostingClassifier	**1.00**	**0.97**	0.16	**1.00**	**0.97**
SupportVectorClassifier	0.99	0.89	0.32	0.99	0.89
DecisionTreeClassifier	**1.00**	0.91	0.28	**1.00**	0.92
GaussianNaiveBayes	0.95	0.85	**0.38**	0.95	0.85

Table 5.6: Performance metrics of classifiers using 21 features

Model name	F1_score		rmse	accuracy_score	
	Train_data	Test_data		Train_data	Test_data
GradientBoostingClassifier	**1.00**	**0.97**	0.16	**1.00**	**0.97**
SupportVectorClassifier	0.98	0.91	0.28	0.98	0.92
DecisionTreeClassifier	0.99	0.90	0.30	0.99	0.90
RandomForestClassifier	0.99	0.89	0.32	0.99	0.89
GaussianNaiveBayes	0.92	0.81	0.43	0.92	0.81
KNeighborsClassifier	0.85	0.76	**0.47**	0.85	0.77

Table 5.5: Performance metrics of classifiers using 30 features

Model	Before Hyper Tune		After Hyper Tune	
	Train data	Test data	Train data	Test data
Support Vector Classifier	0.98	**0.91**	**1.00**	**0.9067**
Random Forest Classifier	**0.99**	0.89	**1.00**	0.8933

Table 5.7: Comparison before and after hyper parameter tuning

Result & Analysis

Validation data accuracy		
Model name	**30 features**	**20 features**
K-Neighbors Classifier	**0.45**	**0.50**
Linear Svm	0.43	0.43
Logistic Regression	0.43	0.47
Random Forest Classifier	0.43	0.39
Radial Svm	0.41	0.41
Gradient Boosting Classifier	0.39	0.39
Decision Tree Classifier	0.39	0.37
Gaussian Naive Bayes	0.395	0.45

Table 5.8: Validation dataset accuracy

06

Conclusion

Conclusion:

- **Limitation:**
  - Inadequate data to feed the model.
  - Could not collect sub parts of each average columns.
  - Demographic fields prevents better accuracy.

- **Future work:**
  - Using emotional data delve into psychological research.
  - Create a model that will work for globally not demographically.

BIBLIOGRAPHICAL REFERENCES

- [1] J. C. Walsh, M. Comar, J. Folan, S. Williams, and S. Kola-Palmer, “The psychological and behavioural correlates of covid-19 vaccine hesitancy and resistance in ireland and the uk,” Acta Psychologica, vol. 225, p. 103550, 2022. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0001691822000658
- [2] C. Rancher, A. D. Moreland, D. W. Smith, V. Cornelison, M. G. Schmidt,J. Boyle, J. Dayton, and D. G. Kilpatrick, “Using the 5c model to understand covid-19 vaccine hesitancy across a national and south carolina sample,” Journal of Psychiatric Research, vol. 160, pp. 180–186, 2023. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0022395623000778
- [3] A. Raut, A. Samad, J. Verma, and P. Kshirsagar, “Acceptance, hesitancy and refusal towards covid-19 vaccination,” Clinical Epidemiology and Global Health, vol. 21, p. 101283, 2023. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S2213398423000702
- [4] H. Gupta and O. P. Verma, “Vaccine hesitancy in the post-vaccination covid-19 era: A machine learning and statistical analysis driven study,” Evolutionary Intelligence, vol. 16, no. 3, pp. 739–757, 2023.
- [5] M. Saifuzzaman, M. M. Rahman, S. F. Shetu, and N. N. Moon, “Covid-19 and bangladesh: situation report, comparative analysis, and case study,” Current Research in Behavioral Sciences, vol. 2, p. 100034, 2021.
- [6]M. Adamus, V. ˇCavojov´a, and E. Ballov´a Mikuˇskov´a, “Emotional drivers of the vaccination hesitancy and refusal: A dataset from slovakia,” Data in Brief, vol. 47, p. 108980, 2023. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S2352340923000987

THANKS!

Do you have any questions?