Exploring the Influence of Perceived Risk, Vaccine Effectiveness, and Doctor Recommendation on Influenza Vaccine Uptake: A Comparative Analysis using Random Forest and XGBoost Classifiers

Ranjith Ramaswamy, Lokeshwaran Arunachalam
psxrr7@nottingham.ac.uk, psxla5@nottingham.ac.uk
University of Nottingham, UK

***Abstract*—This research paper presents an in-depth analysis of the various factors affecting the uptake of influenza vaccination. Utilizing the Random Forest and XGBoost classifiers, we examined a comprehensive dataset sourced from the 2009 National Health Survey. This dataset encompasses a wide range of information, including vaccination coverage, opinions, and other pertinent factors. The study specifically investigated the influence of perceived risk, vaccine effectiveness, and doctor recommendation on vaccination uptake. By harnessing the capabilities of these advanced machine learning models, we gained valuable insights into the interplay of these factors in individuals' decisions to receive the influenza vaccine. The outcomes emphasize the significance of considering factors such as perceived risk and vaccine effectiveness, alongside the impact of doctor recommendations, in promoting vaccine uptake. These insights have the potential to inform evidence-based strategies aimed at enhancing public health outcomes and optimizing vaccination programs.**

***Keywords*—vaccination, Influenza, H1N1 flu, seasonal flu, vaccine uptake, public health, Random Forest algorithm, predictors, public health strategies.**

I. Introduction

Influenza, a widely recognized global health concern, continues to pose significant challenges. The annual outbreaks of flu result in severe illness and even death, emphasizing the urgent need for effective preventive measures. Vaccination has long been acknowledged as a highly effective strategy to reduce influenza-related complications and curb the transmission of the virus within communities. However, despite the availability of vaccines, there exists a persistent issue of inadequate influenza vaccination rates among specific population groups.

To address this critical issue, it is crucial to delve into the intricate factors that shape individuals' decisions regarding influenza vaccination. This study aims to explore the impact of factors like perceived risk, vaccine effectiveness, and doctor recommendation, on the uptake of influenza vaccines. Leveraging a comprehensive dataset obtained from the 2009 National Health Survey, we embark on a journey to unravel the complex dynamics surrounding individuals' vaccination behaviours.

The dataset utilized in this study comprises a wealth of invaluable information, encompassing vaccination coverage data, individual opinions on influenza vaccines and disease, as well as pertinent factors related to healthcare provider recommendations. This extensive dataset, collected during the 2009 influenza season, offers a unique opportunity to meticulously analyse and identify the key determinants that influence individuals' choices to receive the influenza vaccine.
To unravel the intricate relationships within this multifaceted domain, we employ two cutting-edge machine learning models: the Random Forest and XGBoost classifiers. These advanced models possess robust predictive capabilities, enabling us to comprehensively examine the multifaceted interactions among the various factors influencing vaccine uptake. By harnessing the power of these models, we aspire to extract profound insights and pinpoint the most influential factors associated with influenza vaccine uptake.
The outcomes of this study will significantly contribute to our understanding of the underlying factors that drive influenza vaccination uptake. Furthermore, these findings will serve as a cornerstone for developing evidence-based strategies that effectively promote vaccination within specific target populations. Ultimately, enhancing influenza vaccine uptake rates holds the potential to alleviate the burden of influenza-related illnesses and mitigate the associated complications.

II. Literature Review

Several researchers have utilized the National 2009 H1N1 Flu Survey (NHFS) dataset and similar datasets to investigate factors influencing influenza vaccine uptake. This literature review provides an overview of key methods adopted by researchers in analysing this dataset and extracting insights related to vaccine uptake.(1)

Ding, Santibanez, Jamieson, et al. (2011) conducted a national survey to examine influenza vaccination among pregnant women during the 2009-2010 season. Their methodology involved telephone interviews and logistic regression analysis to determine factors associated with vaccination coverage. Results showed higher coverage among pregnant women who received healthcare provider recommendations and perceived vaccines as effective. This highlights the role of healthcare providers and messaging in influencing vaccination behaviour.(2)

In another investigation, Santibanez, Santoli, Bridges, et al. (2013) conducted a comprehensive study during the 2009–2010 influenza season to understand opinions about the 2009 pandemic influenza A (H1N1) and seasonal influenza vaccination among adults. They employed data from the National H1N1 Flu Survey and focused on identifying socio-demographic disparities in vaccination opinions and perceptions of the disease. Multivariable logistic regression analysis was utilised to assess the association between vaccination coverage and factors such as health care provider recommendation, perceived vaccine effectiveness, and perceived risk of influenza infection.(3)

Rubinstein, Marcu, Yardley, et al. (2015) applied the COM-B (capability, opportunity, motivation, and behaviour) model to identify barriers and facilitators influencing postpartum vaccination uptake. Through focus groups and interviews with the public, they conducted thematic analysis to gain insights into the diverse factors influencing vaccine uptake. The COM-B model provided a theoretical framework that allowed for a comprehensive understanding of the various components of behaviour influencing vaccination decisions.(4)

Blackwell (2015) examined factors associated with the receipt of influenza A (H1N1) pdm09 vaccinations among US children. Using data from the National Health Interview Survey, the study explored demographic characteristics, family structure, socioeconomic status, access to health care, and chronic condition status as potential predictors of vaccine uptake. Logistic regression models were employed to estimate the receipt of the first and second vaccine doses within specific timeframes.(5)

In conclusion, the studies reviewed in this analysis have significantly contributed to our understanding of the factors influencing influenza vaccine uptake. Through the utilization of methodologies such as logistic regression analysis, thematic analysis, and theoretical frameworks like the COM-B model, these researchers have provided valuable insights into the determinants of vaccination behaviour. Their findings underscore the importance of health care providers, effective communication strategies, and addressing multiple components of behaviour in promoting vaccine uptake.

III. Methodology

To address the research objectives of this study, a rigorous methodology was implemented, comprising data preprocessing techniques, correlation analysis, statistical tests, and the application of two machine learning models. The selected approaches were justified based on their effectiveness in handling missing values, reducing dimensionality, and exploring the factors influencing influenza vaccine uptake.(6)

A. *Data pre-processing*

The dataset was meticulously examined, and due to notable disparities of attributes relationship between H1N1 and seasonal flu vaccination, the data were processed separately for both type of vaccinations. From the ***Fig.1.*** it is clear that the proportion of vaccination for H1N1 and seasonal flu is not similar for the same age group. This proves the attribute values will be different for H1N1 and seasonal flu vaccination.

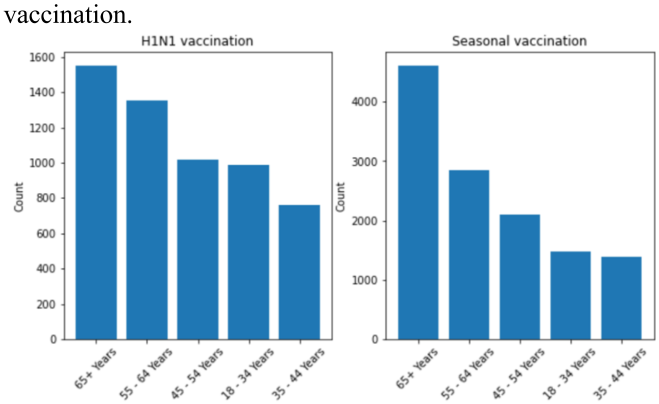

Fig.1. H1N1 and seasonal vaccine proportion based on age group

B. *Managing Missing values*

The missing values for categorical attributes are replaced using proportional imputation. The proportional imputation is a technique which calculates the proportion of all values in a category and compare it with the missing values proportion. The categorical value which has same proportion as missing value is used to impute missing data. From the Fig.2. the missing value of H1N1 flu concern data is represented by -1 and the proportion of missing data is equal to level of concern 1. Therefore, the missing value will be replaced with value 1.

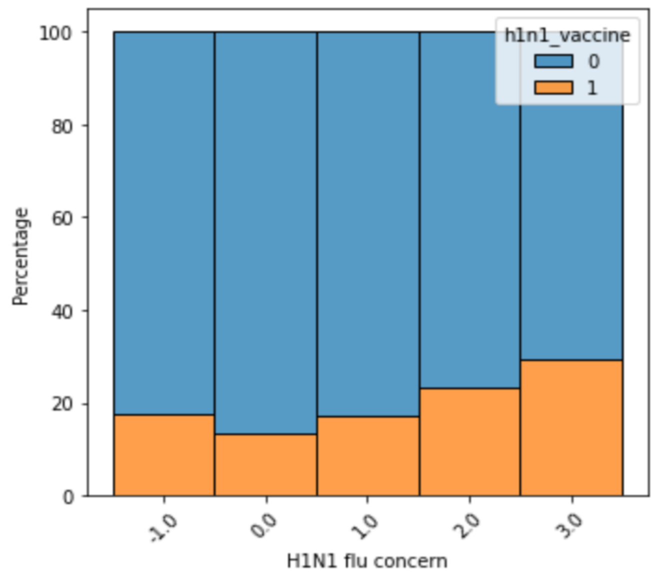

Fig.2. The H1N1 vaccination based on H1N1 concern with missing value

C. *Dimensionality Reduction*

To overcome the challenge posed by columns with numerous unique categorical values, a dimensionality reduction technique was employed. Specifically, columns such as geographical region, employment industry, and employment occupation were transformed using a percentage-based approach. This involved replacing the original values with the percentage of H1N1 and seasonal flu vaccination corresponding to each value in the respective column. This reduction technique facilitated better training of the machine learning algorithms while retaining the critical relationship between variables.

D. *Categorical variable Encoding*

3.1) Ordinal Encoding

Some of the categorical feature values have some inherent order such as age group, education, employment status. Based on the order of the values numerical value were assigned from 1 to k. where k is the number of unique categorical values.

3.2) One-Hot Encoding

The categorical features which have no natural order are encoded using One-Hot encoding. Each unique categorical values are represented as a separate feature. The values of the encoded column will be 1 if that unique categorical value is present and 0 otherwise.

E. Dataset Split

The dataset is divided into training and test set. The percentage of training set is 80% and the test set is 20%. The data is shuffled and then divided to remove any correlation in the dataset order and to achieve randomness.

F. Machine learning model

1. Random Forest classifier

The random forest classifier uses an approach of training many decision trees using the same algorithm but different datasets. The dataset used to train decision trees were sampled with replacement this method is called bagging.

Individual decision trees are called as predictors. The prediction for an instance can be obtained by taking the mode value of the predictors result. Fig.3.

2. XGBoost Classifier

XGBoost classifier uses a technique of gradient boosting. Like Random Forest classifier XGBoost classifier have many predictors. Each predictor is used to fit the residual errors of the previous model. Using XGBoost classifier we can apply some regularization in-order to counter overfitting in the model. Fig.4.

G. Statistical Tests

1. Gini index

The Gini index is used to calculates the probability of a specific feature that is not classified correctly when selected at random. The Gini index value rages from 0 to 1. If the value is low, then we can assume there is high chance to classify correctly.

$$Gini\ Index = 1 - \sum_{i=1}^{n} (P_i)^2$$

The value P_i denotes the probability of an individual data being classified for a particular class.

2. Chi-Square test

The Chi-Square test is used to find whether two categorical attribute values are independent or related to each other.

$$Chi\ squared = \sum (O_i - E_i)^2 / E_i$$

O_i : The observed value
E_i : The Expected value

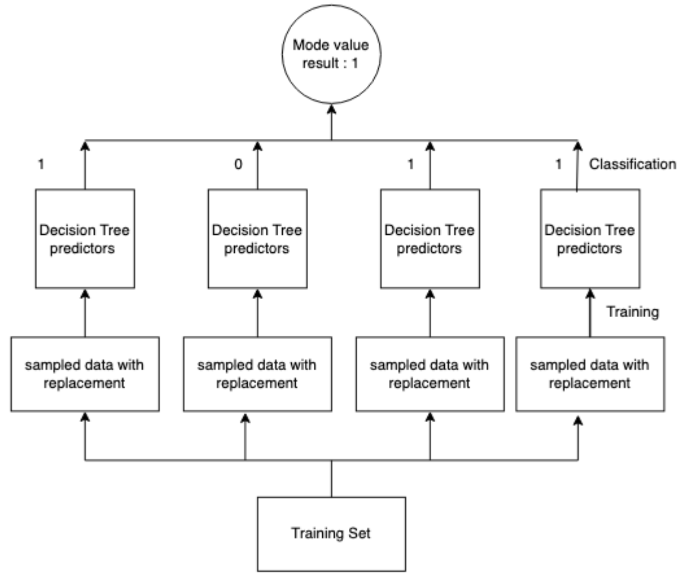

Fig.3. Random Forest Classifier Architecture

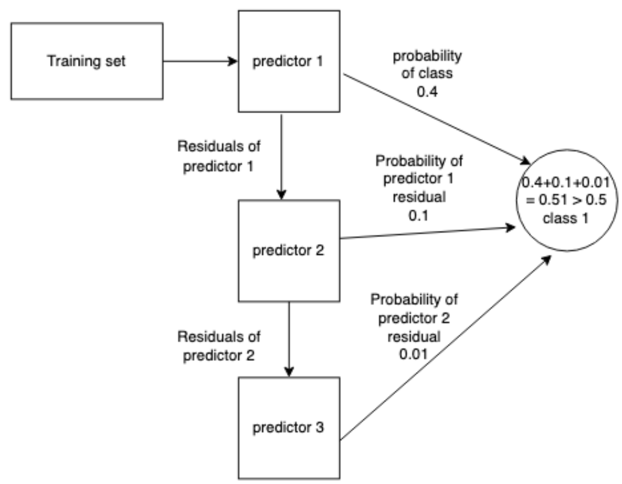

Fig.4. XGBoost Classifier Architecture

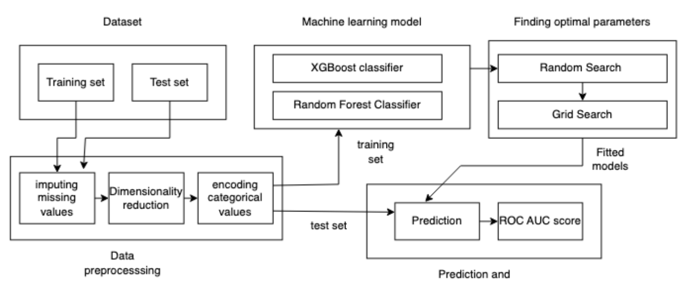

Fig.5. Machine learning method architecture

H. Analysis of Factors and Models:

Correlation analysis was conducted to identify the factors influencing influenza vaccine uptake. The correlations were examined to determine the strength and direction of the relationships between variables. Subsequently, the Random Forest and XGBoost(7) models were employed to further analyse and validate the identified factors. These models, renowned for their predictive capabilities, were fitted to the pre-processed dataset, enabling the exploration of intricate associations within the data. Additionally, statistical tests were employed to ascertain the significance of the identified factors and their impact on vaccine uptake.(8)

The selected approaches in data pre-processing, dimensionality reduction, and analysis of factors and models provided a robust framework for investigating the factors influencing H1N1 and seasonal flu vaccine uptake. By employing this methodology, we aimed to unravel the complexities surrounding vaccine uptake and contribute valuable insights to inform evidence-based strategies for promoting vaccination and mitigating the impact of influenza outbreaks.

IV. Results

A. Data Analysis:

During the analysis phase, we observed significant correlations between vaccine intake and several factors. For H1N1 vaccine, the following factors showed high correlation coefficients: doctor recommendation (0.39), opinions on H1N1 vaccine effectiveness (0.27), and H1N1 risk (0.32) Fig.6. Similarly, for seasonal vaccine, strong correlations were found with doctor recommendation (0.36), opinions on seasonal vaccine effectiveness (0.36), and seasonal risk (0.39) Fig.7. These correlations indicate the influence of these factors on vaccine uptake.

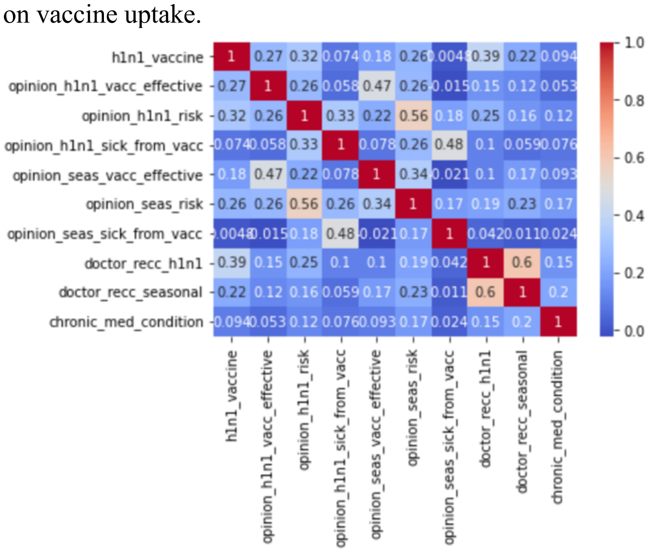

Fig.6. H1N1 vaccine correlation analysis

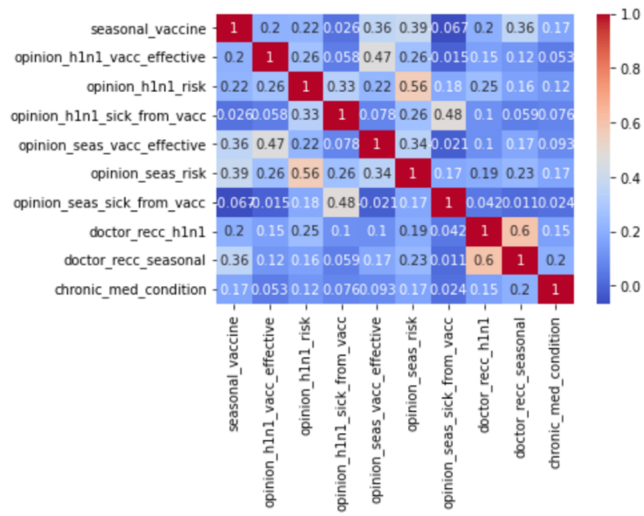

Fig.7. Seasonal Flu vaccine correlation analysis

B. Pre-processing:

To prepare the data for further analysis, we employed specific pre-processing techniques. As the data varied between H1N1 and seasonal flu vaccination, we handled them separately. To address dimensionality, we replaced columns with a high number of categorical values by encoding the percentage of vaccination for each corresponding value. Proportional imputation was employed to handle missing values, which yielded superior results compared to using new unique values like -1 or "unknown". In cases where similar proportions were not available, unique values were used instead. Ordinal encoding was applied to columns with inherent order, while categorical columns were transformed to minimize dimensionality.

C. Classification:

For classification purposes, two models were utilized: XGBoost and Random Forest. The XGBoost model achieved an ROC AUC score of 0.873, 0.858 for H1N1 and Seasonal flu vaccine prediction, while the Random Forest model achieved a score of 0.87, 0.859. Both machine learning model accuracy were similar.

The random forest estimators based on opinion of people regarding vaccination and doctor recommendation revealed how these factors influence the vaccination.(9)

The random forest decision tree estimator analysis revealed that for H1N1 vaccination, individuals who did not receive a doctor recommendation, perceived lower H1N1 risk, and considered the H1N1 vaccine less effective were more likely to be unvaccinated. Conversely, those who received a doctor recommendation and perceived the vaccine as effective were more inclined to be vaccinated. Similar patterns were observed for seasonal vaccination.

D. Statistical Test Results

Additionally, Gini values were examined to assess feature importance in the Random Forest model. For H1N1 vaccination, individuals who lacked a doctor recommendation, had a perceived seasonal vaccine risk value 2.5 or less, and perceived H1N1 vaccine effectiveness value 4.5 or less had a Gini value of 0.38, indicating a higher likelihood of not taking the vaccine. Conversely, individuals who received a doctor recommendation and perceived the seasonal vaccine effectiveness value 2.5 or greater had a Gini value of 0.28, suggesting a higher likelihood of vaccination.

Furthermore, chi-square tests revealed significant relationships between doctor recommendation of H1N1 vaccine and H1N1 vaccination, as well as between opinions on H1N1 vaccine effectiveness and H1N1 vaccination. Similarly, significant relationships were found between opinion on seasonal vaccine effectiveness, doctor recommendation, and seasonal vaccination.

These findings provide valuable insights into the relationships between various factors and influenza vaccination behaviour. The correlations, Gini values, and chi-square tests contribute to our understanding of the key determinants influencing vaccine uptake. This knowledge can guide the development of targeted interventions and strategies aimed at improving vaccination rates and enhancing public health outcomes.

Table 1. H1N1 vaccine prediction evaluation metric scores

Classifier	Precision score	Recall score	F1 Score	ROC AUC Score
XGBoost	0.73	0.5	0.59	0.873
Random Forest	0.68	0.53	0.60	0.87

Table.2. Seasonal flu vaccine prediction evaluation metric scores

Classifier	Precision score	Recall score	F1 Score	ROC AUC Score
XGBoost	0.77	0.75	0.76	0.858
Random Forest	0.76	0.77	0.77	0.859

Table.3. Machine learning Model parameters

Classifier	parameters
H1N1 vaccination	
XGBoost	{"n_estimators":320,"objective":"binary:logistic", "eval_metric":"auc","gamma":0.5,"alpha":1, "subsample":0.7,"max_depth":8, "learning_rate":0.01}
Random Forest	{"n_estimators":360,"min_samples_leaf":2, "max_features":"sqrt","class_weight":"balanced", "max_samples":0.7, "max_depth":28}
Seasonal Flu vaccination	
XGBoost	{"n_estimators":250,"objective"="binary:logistic", "eval_metric":"auc","gamma":0.25,"alpha":0, "subsample":0.9,"max_depth":8, "learning_rate":0.01}
Random Forest	{"n_estimators":366,"min_samples_leaf":4, "max_features":"sqrt","class_weight":"balanced", "max_depth":25}

V. DISCUSSION

In comparing the approaches and findings of the two studies, several similarities and differences can be observed.

A. Data Analysis and Pre-processing

In both studies, separate pre-processing approaches were employed to address variations in vaccination percentages for H1N1 and seasonal flu. These variations reflect the distinct characteristics and public perception associated with each type of influenza. By recognizing the need for separate pre-processing. The studies demonstrate an understanding of the importance of addressing data discrepancies.

One study utilized a method of replacing missing values with specific values (-1 and "unknown"). This approach allowed for explicit identification of missing data points in the dataset. By assigning specific values, the researchers could differentiate missing values from valid responses, ensuring the integrity of the dataset for further analysis. The other study employed proportional imputation to handle missing values. This approach aimed to maintain the same distribution of the data by filling in missing values based on the proportions of the available data. Proportional imputation helps preserve the overall characteristics of the dataset while addressing missing values.

B. Classification and Model Selection

Both studies employed different algorithms for classification tasks, namely XGBoost and Random Forest. These algorithms are popular choices in machine learning due to their ability to handle complex interactions and capture nonlinear relationships in the data. The selection of these algorithms indicates a recognition of their potential benefits in predicting vaccine uptake.

One study used to train the model without balancing the target column values (i.e H1N1 or seasonal Flu vaccination). The other study used balanced target column values which yielded better results than the unbalanced data.

To optimize the performance of the selected models, both studies utilized hyperparameter tuning techniques. The randomized search, which randomly selects combinations of hyperparameters from predefined ranges. This approach helps explore a wide range of hyperparameter combinations efficiently. The grid search, which exhaustively searches through all possible combinations of hyperparameters within predefined ranges. Grid search provides a comprehensive evaluation of various hyperparameter settings.

C. Model Evaluation and Performance Metrics

Evaluation of the models was conducted using various performance metrics, including precision, recall, and F1 scores. These metrics assess the model's ability to correctly classify positive and negative instances, considering both the true positives and false positives.

One study placed emphasis on maximizing the area under the curve (AUC) score as the primary evaluation metric. A higher AUC score indicates better overall performance of the model.

The other study employed AUC score using K-fold cross-validation to estimate the performance of the models. Cross-validation is a resampling technique that divides the data into Subsets of K Folds. Each fold is used as to validate the trained model for K times. The remaining data is used to train the model. Averaging the results across multiple iterations, K-fold cross-validation gives a better result of the model's performance.(10)

The AUC score using K-fold cross validation gave better results.

D. Visualization and Interpretation

Both studies utilized visualization techniques to interpret their models and gain insights into the factors influencing vaccination behaviour. Visualizations are powerful tools for understanding complex models and identifying patterns in the data.

One study employed ROC curves to visualize the performance of the models. ROC curves illustrate the trade-off between the true positive rate and the false positive rate at various classification thresholds. This visualization allows for an intuitive understanding of the model's performance across different decision boundaries, providing insights into the models discriminatory power.Fig.8. and Fig.9.

The other study focused on visualizing the first estimator of the Random Forest model using attributes of opinion and doctor recommendation. Random Forest use bagging technique to train decision trees. By visualizing the first estimator, researchers can gain insights into the decision-making process of the Random Forest algorithm, such as the most influential features and their contributions to the classification outcome. This visualization provides valuable information about the factors driving vaccine uptake.Fig.10. and Fig.11.

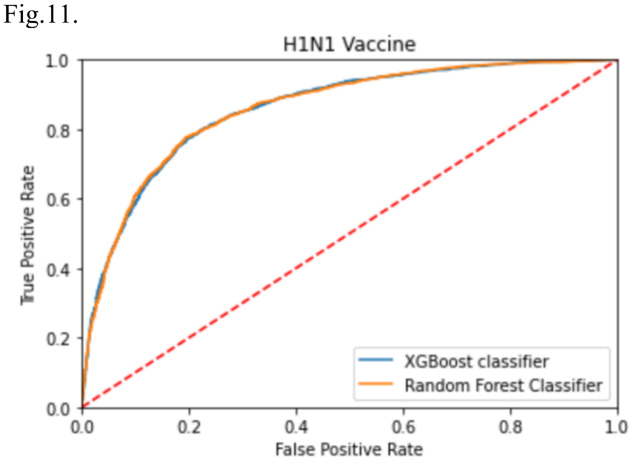

Fig.8. ROC curve for H1N1 vaccine

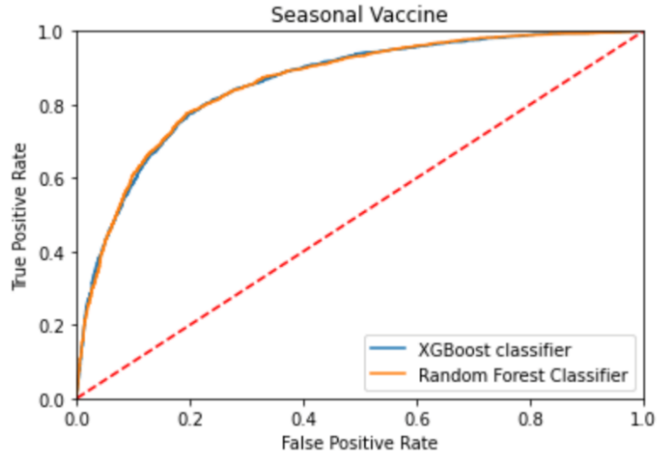

Fig.9. ROC curve for Seasonal vaccine

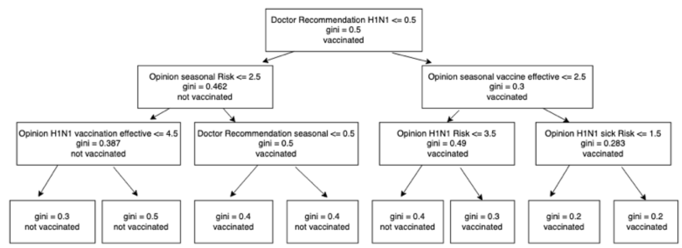

Fig.10. H1N1 vaccination prediction using random forest first estimator with maximum depth 3 considering only people opinion and doctor recommendation on vaccine.

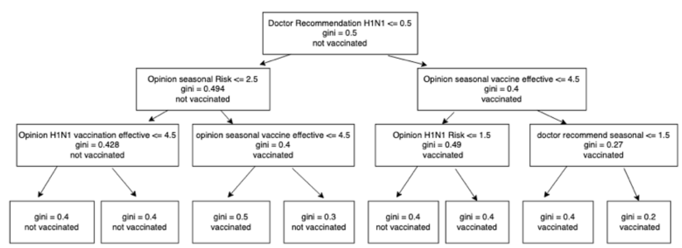

Fig.11. Seasonal flu vaccination prediction using random forest first estimator with maximum depth 3 considering only people opinion and doctor recommendation on vaccine.

E. Correlation Analysis and Significance Testing:

Both studies recognized the importance of exploring the relationships between attributes and vaccine uptake through correlation analysis and significance testing. These analytical approaches provide valuable insights into the factors influencing vaccination behaviour.

One study employed a correlation heat map to visualize attribute correlations. The heat map displayed a color-coded matrix where each cell represented the correlation coefficient between two attributes. This visualization technique allowed researchers to quickly identify strong positive or negative correlations between variables. By examining attribute correlations, researchers gained insights into potential associations with vaccine uptake and identified variables that were highly correlated with vaccination behaviour.Fig.6. and Fig.7.

The other study conducted chi-square tests to examine the relationships between categorical attributes and vaccine uptake. Chi-square tests are widely used to determine whether there is a significant association between two categorical variables. In this study, researchers applied chi-square tests to assess the independence between each attribute and vaccine uptake. By calculating the chi-square statistic and comparing it to the critical value, researchers determined whether the association between the attribute and vaccine uptake was statistically significant. This approach enabled the identification of attributes that were significantly associated with vaccine uptake, highlighting the factors that may have a substantial impact on vaccination behaviour.

By incorporating correlation analysis and significance testing, both studies provided a deeper understanding of the relationships between attributes and vaccine uptake. These analyses allowed researchers to identify significant associations and dependencies, providing valuable insights into the factors influencing vaccination behaviour.

Overall, both studies demonstrated a comprehensive understanding of data analysis, pre-processing, model selection, evaluation metrics, visualization, and significance testing in the context of influenza vaccine uptake. While there were differences in specific techniques and metrics employed, these variations contributed to a broader understanding of the factors influencing vaccination behaviour. By combining the strengths of both approaches, future research can further enhance our understanding of vaccine uptake and develop targeted interventions to improve public health outcomes.

F. Comparison with other research methods:

In comparing our methods with the literature review, similarities and differences emerged. Like the Some of the reviewed studies specifically targeted subpopulations such as pregnant women and children. For instance, Ding, Santibanez, Jamieson, et al. (2011) focused on pregnant women and assessed factors associated with vaccination coverage in this group. Similarly, Santibanez, Santoli, Bridges, et al. (2013) analyzed the NHFS data to understand vaccination opinions among adults, including pregnant women and parents of young children.

In comparison, our study utilized the NHFS dataset to investigate factors influencing influenza vaccine uptake across a broader population. While we did not specifically target pregnant women and children, our findings align with previous research regarding the importance of healthcare provider recommendations and perceptions of vaccine effectiveness.

And like reviewed studies, we focused on identifying factors influencing vaccine uptake, emphasizing healthcare provider recommendations and perceptions of vaccine effectiveness. However, unlike previous studies that relied on logistic regression, we incorporated machine learning algorithms (XGBoost and Random Forest classifier) to improve predictive accuracy. Our study showcased the potential for enhanced performance through machine learning, albeit at the expense of some interpretability. While machine learning models provide higher accuracy, it is crucial to consider the trade-off between predictive performance and interpretability. In contrast, logistic regression analysis offers interpretable results but may not capture complex relationships as effectively. By combining statistical analysis and machine learning, we contributed valuable insights into influenza vaccine uptake determinants. Findings emphasized the importance of healthcare provider recommendations, effective communication strategies, and understanding behaviour components. Our approach demonstrated higher predictive accuracy compared to traditional logistic regression.

VI. Conclusion

In this research paper, we delved into the complex dynamics surrounding influenza vaccine uptake by utilizing advanced machine learning models, specifically the Random Forest and XGBoost classifiers. Our comprehensive analysis of the 2009 National Health Survey dataset provided valuable insights into the factors influencing individuals' decisions to receive the influenza vaccine.

The findings of our study underscore the significance of factors such as doctor recommendation, perceived risk, and vaccine effectiveness in shaping vaccination behaviour. We observed strong correlations between vaccine uptake and these factors, with individuals who received a doctor recommendation and held positive perceptions of vaccine effectiveness being more likely to get vaccinated. Conversely, those lacking a doctor recommendation or perceiving lower risk or effectiveness were less inclined to receive the vaccine.

The application of Random Forest and XGBoost models facilitated a deeper understanding of the multifaceted interactions among these factors. The random forest estimator tree analysis further reinforced the importance of doctor recommendation, perceived risk, and vaccine effectiveness in influencing vaccination behaviour.

VII. Recommendation for future research

Building upon the findings of this study, there are several avenues for future research that can advance our understanding of influenza vaccine uptake and inform evidence-based strategies:

A. Longitudinal Studies:

Conducting longitudinal studies would enable tracking vaccination behaviour over time and assessing the long-term impact of factors such as doctor recommendation, perceived risk, and vaccine effectiveness. This would provide a more comprehensive understanding of the dynamic nature of vaccination behaviour and the persistence of its determinants.

B. Targeted Interventions:

Tailor interventions based on specific population groups and their unique determinants of vaccine uptake. Future research can focus on understanding the factors that influence vaccination behaviour among pregnant women, children, and other high-risk populations. Developing targeted interventions can effectively address the specific barriers and motivations within these groups.

C. Cultural and Socioeconomic Factors:

Investigate the influence of cultural and socioeconomic factors on vaccine uptake. Understanding how cultural beliefs, social norms, and economic barriers impact vaccination behaviour can aid in designing interventions and communication strategies that are culturally sensitive and address specific socioeconomic challenges.

D. Communication Strategies:

Explore the effectiveness of different communication strategies in promoting vaccine uptake. Investigate the impact of various messaging approaches, including educational campaigns, personalized recommendations, and community-based interventions. Identifying the most effective methods for conveying vaccine-related information can significantly improve vaccination rates.

E. Vaccine Hesitancy:

Examining vaccine hesitancy is a crucial determinant of vaccine uptake. Investigate the underlying reasons for vaccine hesitancy and develop strategies to address misinformation, misconceptions, and concerns about vaccine safety and efficacy. Targeted efforts to address vaccine hesitancy can play a pivotal role in increasing vaccine acceptance.

F. Comparative Analysis:

Conduct comparative analyses across different countries or regions to identify variations in vaccine uptake and the factors influencing it. Comparing vaccination behaviours and determinants across diverse populations can provide insights into cultural, social, and policy-related factors that contribute to vaccine acceptance. Such comparative analyses can help develop context-specific strategies to enhance vaccine uptake.

By pursuing these research directions, we can further enhance our understanding of influenza vaccine uptake and contribute to the development of evidence-based strategies for improving vaccination rates and public health outcomes. The insights gained from such research will play a vital role in mitigating the burden of influenza-related illnesses and ensuring the effectiveness of vaccination programs worldwide.

Contribution

Lokeshwaran Arunachalam: proportional imputation, ordinal encoding, training random forest classifier, Gini values using random forest estimator, using cross validation, random search, grid search, and evaluation scores for random forest classifier.

Ranjith Ramaswamy: Dimensionality reduction, One-hot encoding, training XGBoost classifier, correlation heat map, chi-square test, random search, grid search, and evaluation score for XGBoost classifier

VIII. References

1. Zhao AR, Bishai DM. Public health spending, primary care, and perceived risk promoted vaccination against H1N1. Vaccine. 2022;40(2):325–33.
2. Ding H, Santibanez TA, Jamieson DJ, Weinbaum CM, Euler GL, Grohskopf LA, et al. Influenza vaccination coverage among pregnant women–National 2009 H1N1 Flu Survey (NHFS). Am J Obstet Gynecol. 2011;204(6):S96–106.
3. Santibanez TA, Singleton JA, Santibanez SS, Wortley P, Bell BP. Socio-demographic differences in opinions about 2009 pandemic influenza A (H1N1) and seasonal influenza vaccination and disease among adults during the 2009–2010 influenza season. Influenza Other Respir Viruses. 2013;7(3):383–92.
4. Rubinstein H, Marcu A, Yardley L, Michie S. Public preferences for vaccination and antiviral medicines under different pandemic flu outbreak scenarios. BMC Public Health. 2015;15(1):190–190.
5. Blackwell DL. Modeling Receipt of Influenza A(H1N1)pdm09 Vaccinations Among US Children During the 2009–2010 Flu Season. Med Care. 2015;53(2):191–8.
6. Nokeri TC. Data science solutions with Python : fast and scalable models using Keras, Pyspark Mllib, H2O, XGBoost, and scikit-Learn. Apress; 2022.
7. Ester M, Kriegel HP, Xu X. XGBoost: A scalable tree boosting system. In Proceedings of the 22Nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (vol , pg 785, 2016). Geogr Anal. 2022;
8. Machine Learning in Tribology. Basel: MDPI - Multidisciplinary Digital Publishing Institute; 2022.
9. Hutchison D, Kanade T, Kittler J. Multiple Classifier Systems. Berlin/Heidelberg: Springer Berlin / Heidelberg; 2007. (Lecture Notes in Computer Science; vol. 4472).
10. Wong TT, Yeh PY. Reliable Accuracy Estimates from k-Fold Cross Validation. IEEE Trans Knowl Data Eng. 2020;32(8):1586–94.