



Life Threatening Prediction due to the adverse

effects of Covid-19 Vaccination

*Domain:* ***Healthcare Analytics***| *Group*: ***5*** | *Batch:* ***PGPDSE-FT Online Jan 22***

|  |  |
| --- | --- |
| **Batch Details** | ***PGPDSE-FT Online Jan 22*** |
| **TEAM MEMBERS** | 1. **Reshma Rajan** 2. **Hazel Abraham** 3. **Sayan Biswas** 4. **Anoushka Roy** 5. **Krutika Agrawal** |
| **DOMAIN OF PROJECT** | **HEALTHCARE ANALYSIS** |
| **PROPOSED PROJECT TITLE** | Life Threatening Prediction due to the adverse  effects of Covid-19 Vaccination |
| **GROUP NUMBER** | **5** |
| **TEAM LEADER** | **Reshma Rajan** |
| **MENTOR NAME** | **Vikash Chandra** |

|  |  |  |
| --- | --- | --- |
| S.NO | TITLE | PAGE NO |
| 1. | Industry Review | 5 |
| 2. | Literature Review | 6 |
| 3. | Business Objective | 6 |
| 4. | Dataset Review | 7-10 |
| 5. | Univariate Analysis | 11-16 |
| 6. | Bivariate Analysis | 16-22 |
| 7. | Multivariate Analysis | 22-25 |
| 8. | Feature Engineering | 25-26 |
| 9. | Feature Engineering (Class Imbalance and its treatment) | 26 |
| 10. | Encoding | 26-27 |
| 11. | Data preparation and model building | 27-31 |
| 12. | Base model building | 27-31 |
| 13. | Decision Tree | 31-32 |
| 14. | Random Forest | 32-34 |
| 15. | Gradient Boosting | 34-37 |
| 16. | Ada Boosting | 38 |
| 17. | Conclusion | 39 |



***INDUSTRY REVIEW***

In the dynamic digital era, a mix of science, technology, and medicine has revealed new data systems to enhance statistics, healthcare and drug distribution, and health information reporting on clinical judgments.

In the following aspects, data science in health care has made the most recent and rapid progress:

* Using big data with a mix of vast and complex data sets includes electronic medical records, social media, genetic information, and wireless health device data.
* Machine learning and artificial intelligence, for example, can enhance systematic and unstructured data processing.

It is now feasible to detect illness signs at an early stage because to the application of Data Science in healthcare. Doctors can also monitor patients' status from remote locations thanks to the development of different modern instruments and technology.

***Predictive Analytics in Healthcare***

Predictive analytics is significant in the healthcare industry. It ranks among the most well- liked subjects in health analytics. A predictive model makes accurate forecasts by using previous data, learning from it, and identifying patterns.

It identifies diverse symptom connections and associations, uncovers behaviours and diseases, and then formulates insightful predictions.

Enhancing patient care, managing chronic diseases, and boosting the effectiveness of supply chains and pharmaceutical logistics are all made possible by predictive analytics.

Predictive analytics is seeing increased interest in the field of population health management. It is a data-driven strategy with a focus on disorders that are frequently prevalent in society.

With the use of data analytics, hospitals are able to anticipate patient health declines, offer preventative care, and initiate early therapy to help lower the likelihood of further deterioration.

Furthermore, the logistical supply of hospitals and pharmaceutical departments is closely monitored thanks to predictive analytics.

### The key benefits of predictive analytics in healthcare are listed below:

Helping to manage chronic diseases.

Efficient monitoring and analysis of pharmaceutical logistics demand. Predicts a patient's condition and suggests preventive actions.

Provides faster documentation of hospital data. Predict future medical crises for a patient.



***LITERATURE REVIEW***

The COVID-19 pandemic affected people of all ages around the world. Therefore, this vaccine was developed and made publicly available in an unprecedented era. The Vaccine Adverse Event Reporting System; is a national early warning system for detecting potential safety issues with vaccines approved in the United States.

VAERS is a passive reporting system. That is, individuals submit a report of their experience to the CDC and FDA. VAERS is not designed to determine if a vaccine has caused health problems, but it has abnormal or unexpected patterns of adverse events that may indicate potential vaccine safety problems. Especially useful for detecting. In this way, VAERS can provide the CDC and FDA with valuable information that additional work and evaluation is required to further assess potential safety concerns.

The main purposes of VAERS are:

Detect new, abnormal or rare adverse events of the vaccine. Monitor the increase of known adverse events.

Identify potential patient risk factors for certain types of adverse events. Assess the safety of the newly approved vaccine.

Identify and address potential reporting clusters (for example, Time or Geographical or Product / Batch / Batch Specific Adverse Event Reports that appear to be localized).

Detects persistent safe use issues and management errors.

## Business Objective:

* 1. The goal of this project is to describe life threatening events after COVID-19 vaccination in patients.
  2. In addition, we are going to identify factors associated with more severe adverse effects.

## Approach:

1. Data Understanding
2. Data Pre-Processing
3. Exploratory Data Analysis
4. Data Mining
5. Model Building
6. Model Evaluation
7. Model Optimization

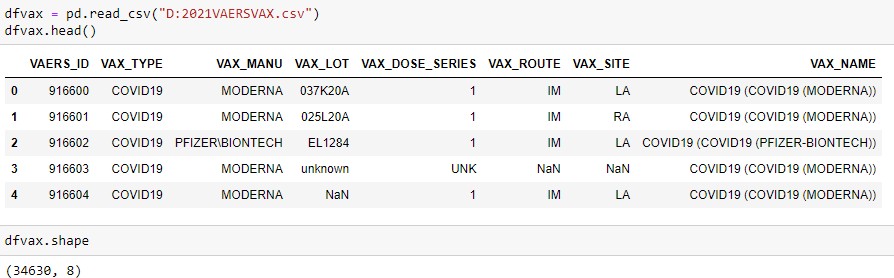
This is a Healthcare domain in which we are trying to predict as what is the possibility of death post vaccination. There are 49021 records and there are 51 features and 1 column(died) which contains the information of what is the possibility of death post vaccination.

We have taken 3 datasets -vaccine dataset, symptom dataset, data dataset

1. **Vaccine dataset**- Provides vaccine information (e.g., vaccine name, manufacturer, lot number, route, site, and number of previous doses administered)

|  |  |  |
| --- | --- | --- |
| VAERS\_ID | Num | VAERS Identification Number |
| VAX\_TYPE | Char | Administered Vaccine Type |
| VAX\_MANU | Char | Vaccine Manufacturer |
| VAX\_LOT | Char | Manufacturer's Vaccine Lot |
| VAX\_DOSE\_SERIES | Char | Number of doses administered |
| VAX\_ROUTE | Char | Vaccination Route |
| VAX\_SITE | Char | Vaccination Site |
| VAX\_NAME | Char | Vaccination Name |

# #Reading csv file-

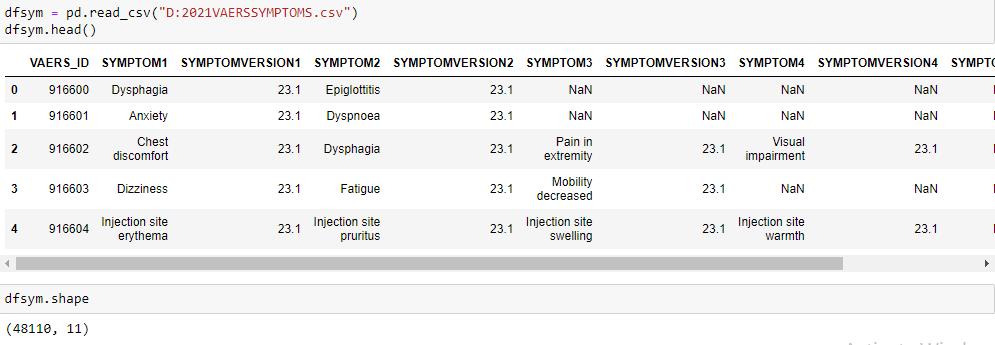


1. **Symptom dataset**- Provides the adverse event coded terms utilizing the MedDRA dictionary.

|  |  |  |
| --- | --- | --- |
| VAERS\_ID | Num | VAERS Identification Number |
| SYMPTOM1 | Char | Adverse Event MedDRA Term 1 |
| SYMPTOMVERSION1 | Num | MedDRA dictionary version number 1 |
| SYMPTOM2 | Char | Adverse Event MedDRA Term 1 |
| SYMPTOMVERSION2 | Num | MedDRA dictionary version number 2 |
| SYMPTOM3 | Char | Adverse Event MedDRA Term 3 |
| SYMPTOMVERSION3 | Num | MedDRA dictionary version number 3 |
| SYMPTOM4 | Char | Adverse Event MedDRA Term 4 |

|  |  |  |
| --- | --- | --- |
| SYMPTOMVERSION4 | Num | MedDRA dictionary version number 4 |
| SYMPTOM5 | Char | Adverse Event MedDRA Term 5 |
| SYMPTOMVERSION5 | Num | MedDRA dictionary version number 5 |

## #Reading csv file-

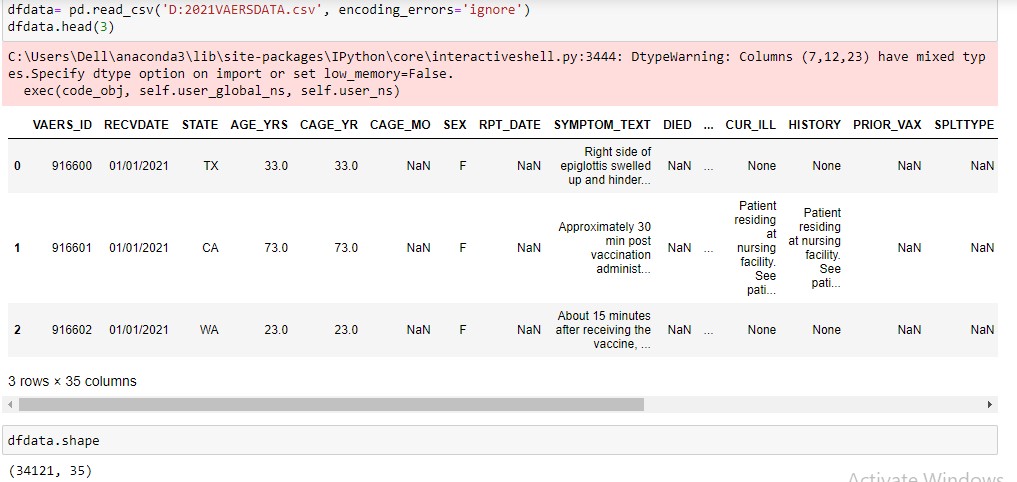


1. **Data dataset**- This dataset contains data like gender, age, and other useful features.

|  |  |  |
| --- | --- | --- |
| VAERS\_ID | Num | VAERS Identification Number |
| RECVDATE | Date | Date report was received |
| STATE | Char | State |
| AGE\_YRS | Num | Age in Years |
| CAGE\_YR | Num | Calculated age of patient in years |
| CAGE\_MO | Num | Calculated age of patient in months |
| SEX | Char | Sex |
| RPT\_DATE | Date | Date Form Completed |
| SYMPTOM\_TE XT | Char | Reported symptom text |
| DIED | Char | Died |
| DATEDIED | Date | Date of Death |
| L\_THREAT | Char | Life-Threatening Illness |
| ER\_VISIT | Char | Emergency Room or  Doctor Visit |
| HOSPITAL | Char | Hospitalized |
| HOSPDAYS | Num | Number of days Hospitalized |
| X\_STAY | Char | Prolongation of Existing Hospitalization |
| DISABLE | Char | Disability |

|  |  |  |
| --- | --- | --- |
| RECOVD | Char | Recovered |
| VAX\_DATE | Date | Vaccination Date |
| ONSET\_DATE | Date | Adverse Event Onset  Date |
| NUMDAYS | Num | Number of days (Onset date - Vax. Date) |
| LAB\_DA TA | Char | Diagnostic laboratory data |
| V\_ADMINBY | Char | Type of facility where  vaccine was administered |
| V\_FUNDBY | Char | Type of funds used topurchase  vaccines |
| OTHER\_MEDS | Char | Other Medications |
| CUR\_ILL | Char | Illnesses at time of  vaccination |
| HISTORY | Char | Chronic or long- standinghealth  conditions |
| PRIOR\_VAX | Char | Prior Vaccination Event  information |
| SPLTTYPE | Char | Manufacturer/Immuniza tion Project Report Number |
| FORM\_VERS | Num | VAERS form version 1 or 2 |
| TODAYS\_DAT E | Date | Date Form Completed |
| BIRTH\_DEFECT | Char | Congenital anomaly or birth defect |
| OFC\_VISIT | Char | Doctor or other healthcare provider  office/clinic visit |
| ER\_ED\_VISIT | Char | Emergency  room/department or urgent care |
| ALLERGIES | Char | Allergies to medications,food, or other products |

## #Reading csv file-



**# Output variable (desired target):**

|  |  |  |
| --- | --- | --- |
| DIED | Char | Yes/No |

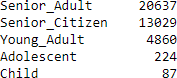
* The output variable named "y" is string which has values either "yes" or "no". Thus we replaced it with binary values: 1 for ‘yes’ and 0 for ‘no’.

The number of **categorical columns** is 34. The number of **numerical columns** is 9.

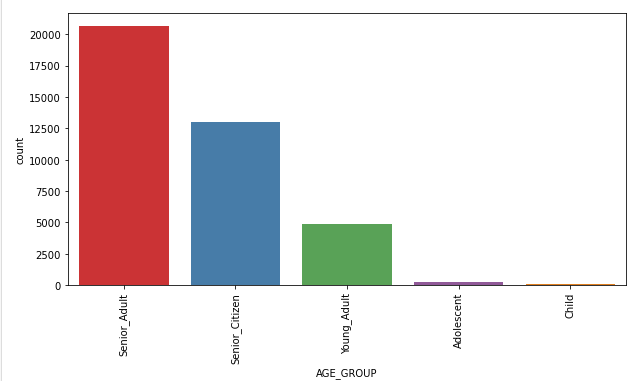
**EXPLORATORY DATA ANALYSIS:**

# Univariate Analysis:

Categorical

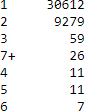
The value counts and pictorial representation of categorical columns : VARIABLE: **AGE\_YRS**

REPRESENTATION OF AGE GROUP:

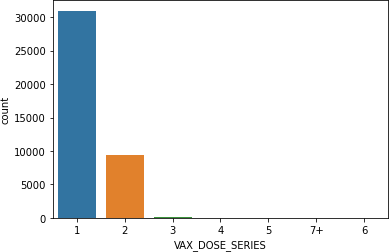


Most of the individuals in our dataset are Senior Adults whose age lies between 30 and 59 followed by Senior\_citizen whose age is greater than 59.

VARIABLE: **VAX\_DOSE\_SERIES**

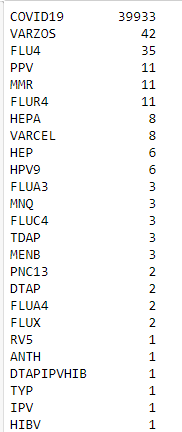


**REPRESENTATION OF VAX\_DOSE\_SERIES**

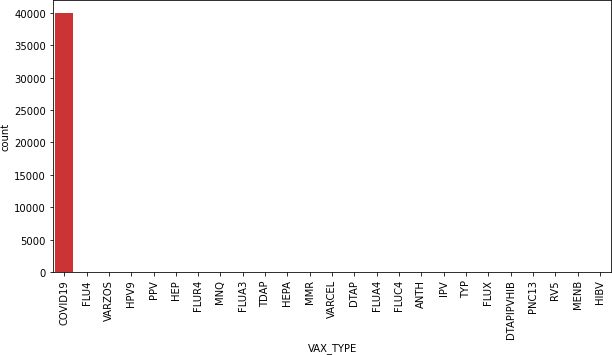


From this countplot , we can see that most of the individuals have taken only 1 vaccine dose.

VARIABLE: **VAX\_TYPE**

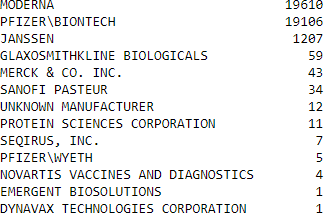


### REPRESENTATION OF VAX\_TYPE

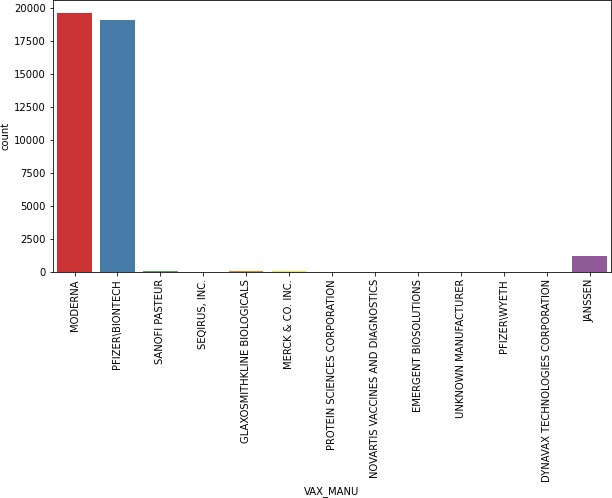


From this plot we can infer that , most of the patients have taken COVID-19 vaccines.

VARIABLE : **VAX\_MANU**



### REPRESENTATION OF VAX\_MANU

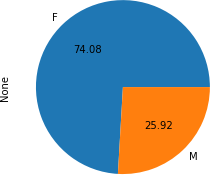


In the above plot, we can clearly see that most of the individuals have taken covid-19 MODERNA vaccie followed by PFIZER.

VARIABLE: **SEX**



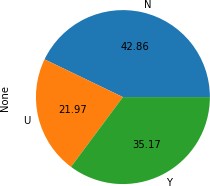
### PICTORIAL REPRESENTATION OF SEX COLUMN



VARIABLE: **RECOVD**



### PICTORIAL REPRESENTATION RECOVD

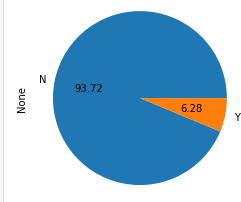


We can see from the above plot that maximum patients did not get recovered post vaccination.

TARGET VARIABLE: **DIED**



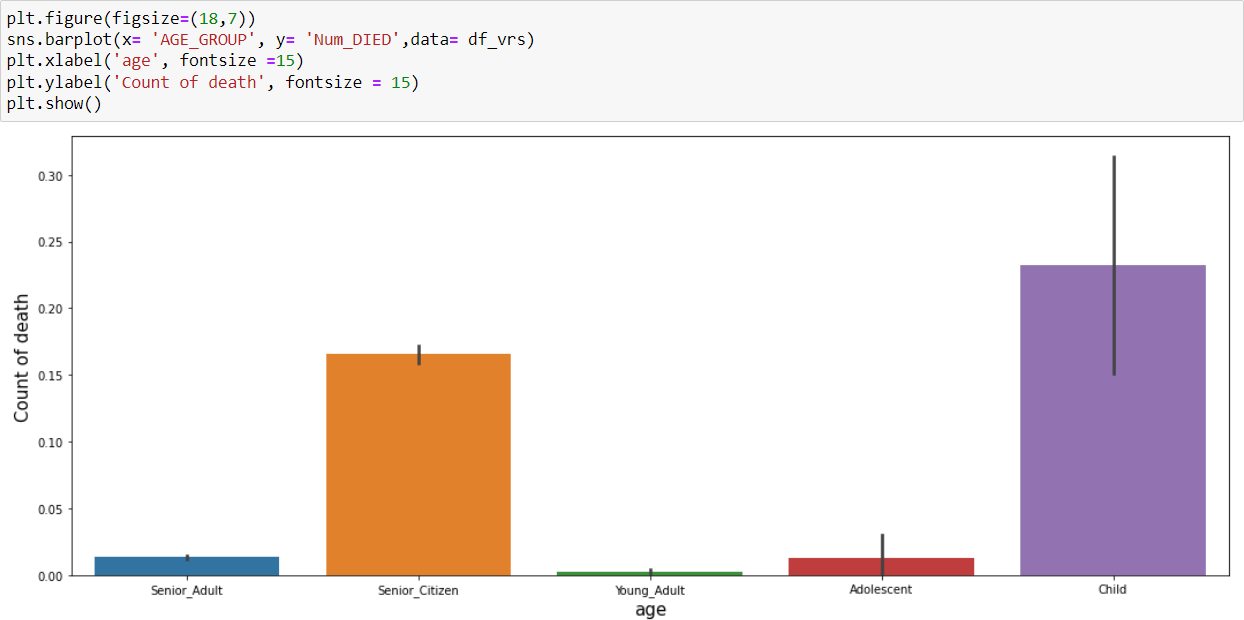
### PICTORIAL REPRESENTATION DIED



As we can see that 93.72% of the pateints have not died post vaccination.

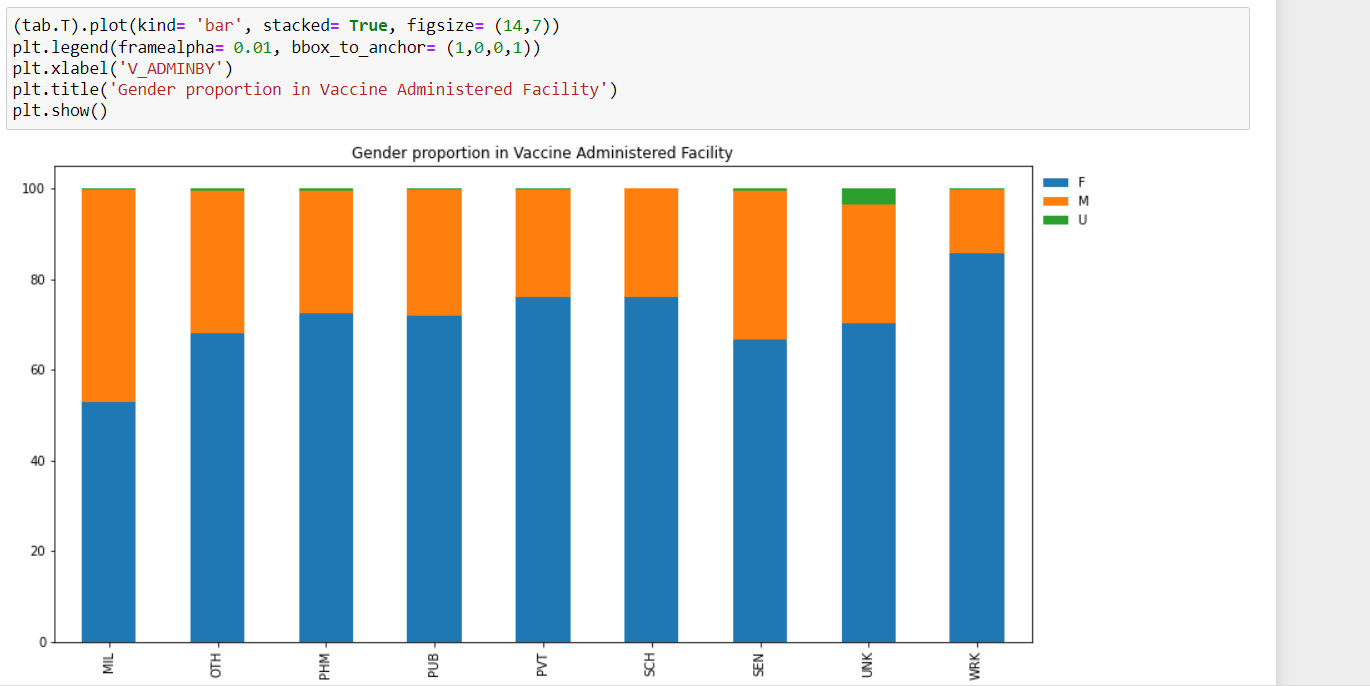
### BIVARIATE ANALYSIS

***PICTORIAL REPRESENTATION OF AGE AND COUNT\_OF\_DEATH***

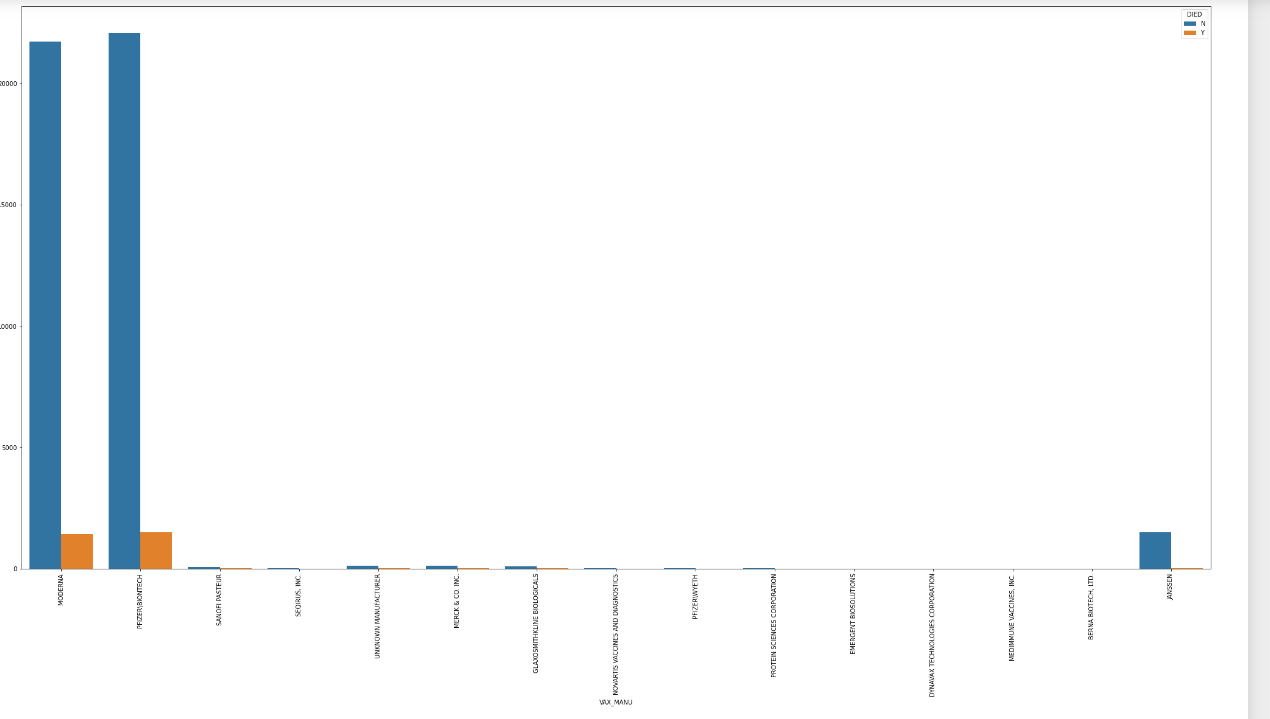


From the above plot we can visualize that most of the death has happened from child age group.

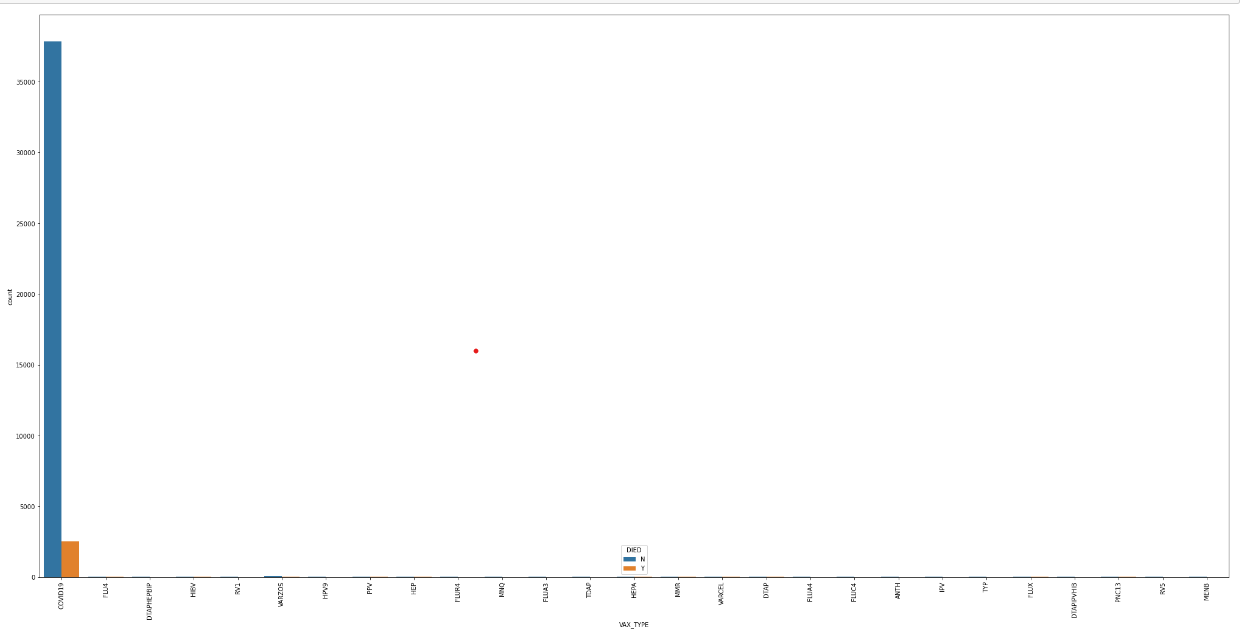
In this Graph Most of the females (85%) have the vaccine dose in their Workplace Clinic (WRK) followed by Private. For the males most of the vaccine is Administered at Military (48%) and the least place where it is Administered is Workplace clinic (WRK)



* + In this Graph Most of the females (85%) have the vaccine dose in their Workplace Clinic (WRK) followed by Private. For the males most of the vaccine is Administered at Military (48%) and the least place where it is Administered is Workplace clinic(WRK)

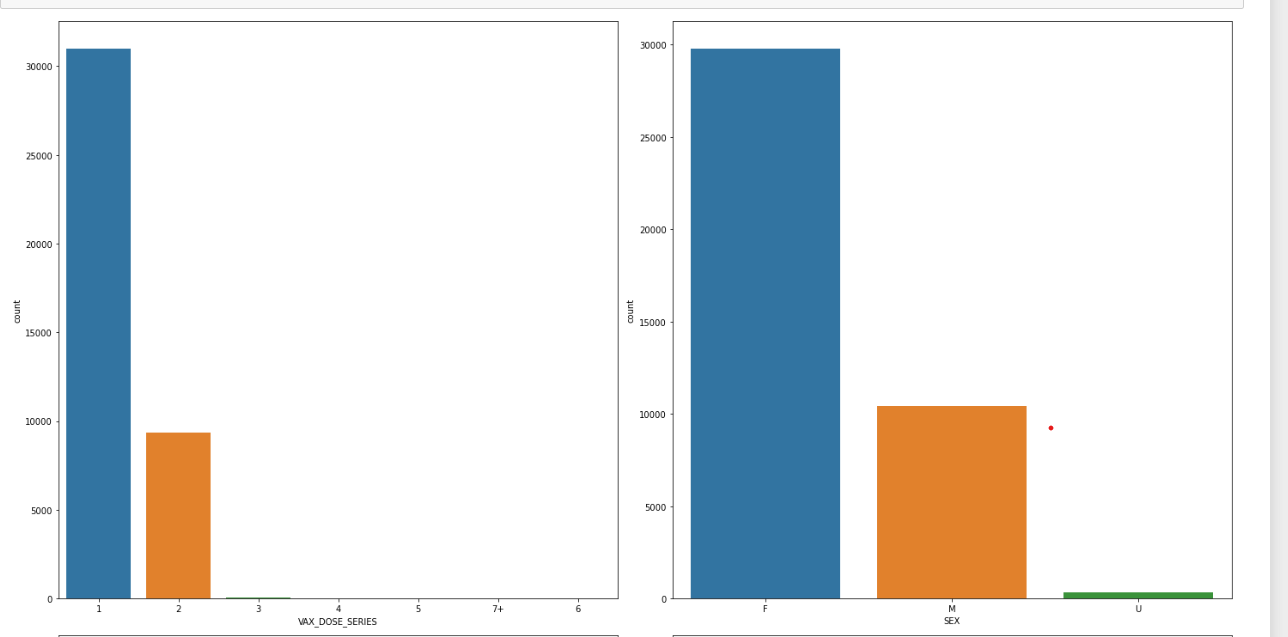


* + This is in Bivariant graph Most of the people have taken Pfizer(Covid -19 ) where the survival rate is high followed by Moderna and Janssen.



* + For this graph we come to know that since the covid 19 vaccine is mostly used its shows that mostly 98% of the people have survived after taking the vaccine. While the 2 % have not survived.

### PLOTS BETWEEN DOSE RECEIVED AND ADMINISTERED BY

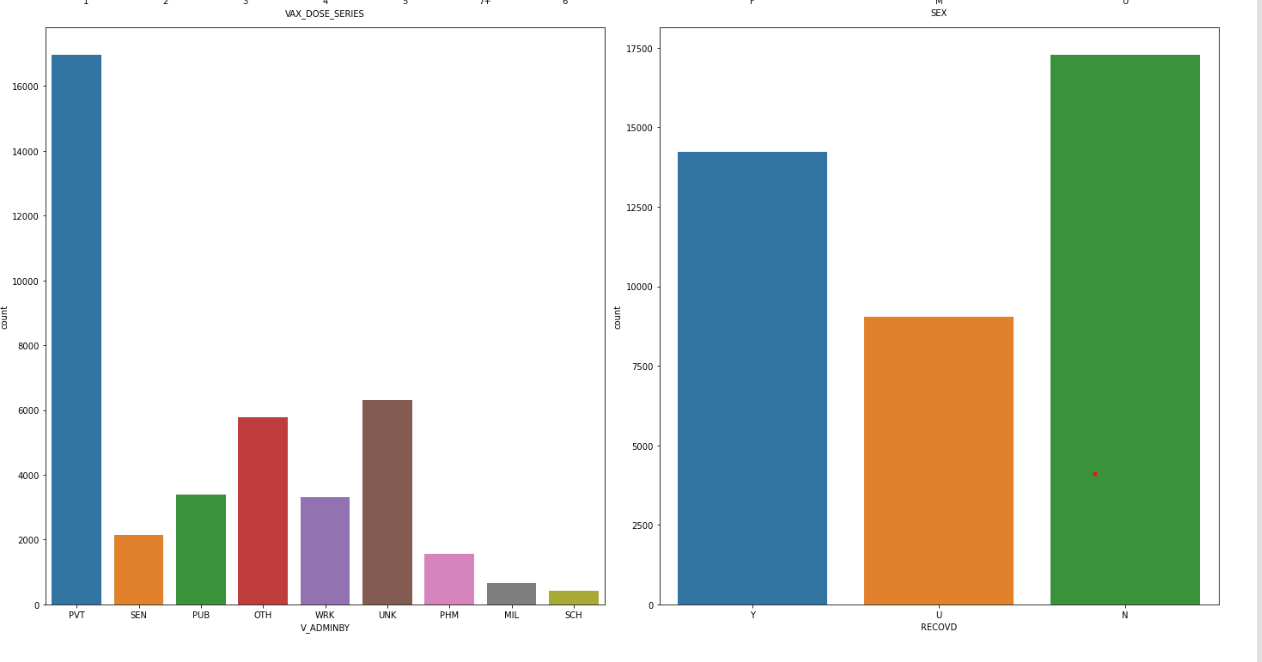


**V\_ADMIN BY**

From this plot we can conclude that many of the vaccine doses are given at the Private places.

**SCH(School)** place shows the least number vaccine given.

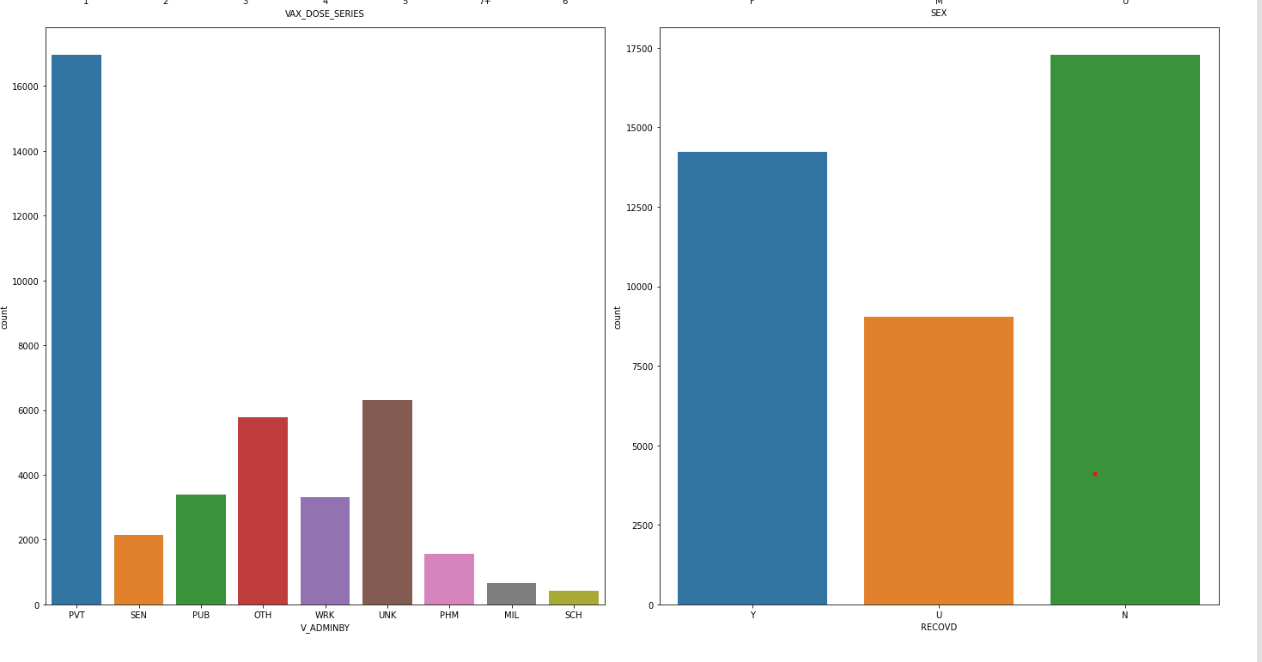
**Vax\_ Doses**



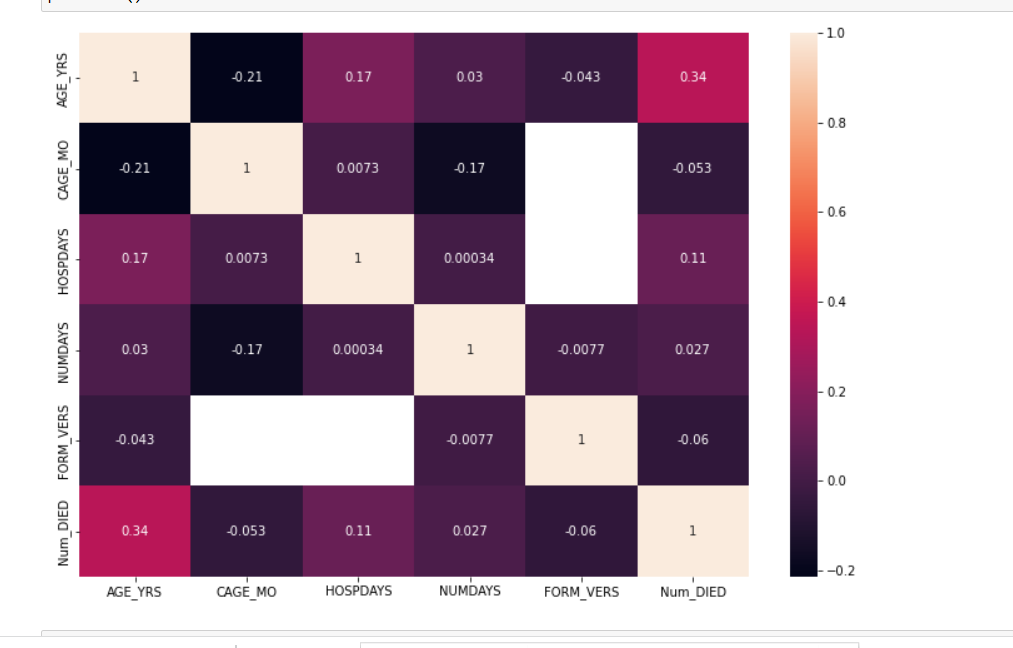
Most of the people have taken only 1 dose for all the vaccine.

**RECOVD**

This plot shows the high count of people who have not yet recovered.



### HEATMAP OF THE DATA



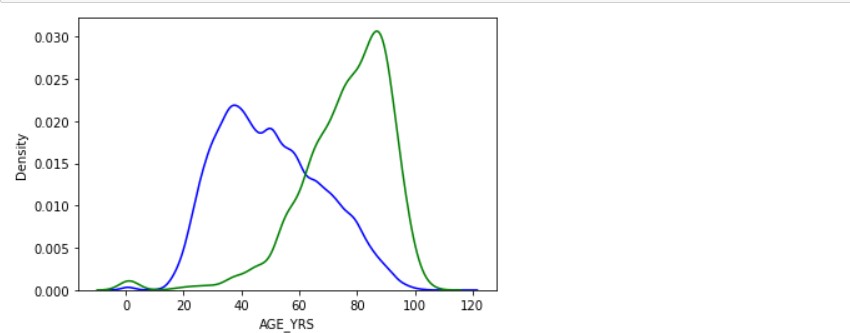
* + 'CAGE\_MO' and 'FORM\_VERS' does not have any correlation between them.
  + 'HOSPDAYS' and 'FORM\_VERS' does not have any correlation between them.
  + Other variables have correlation between them.

**Correlation Plot-**

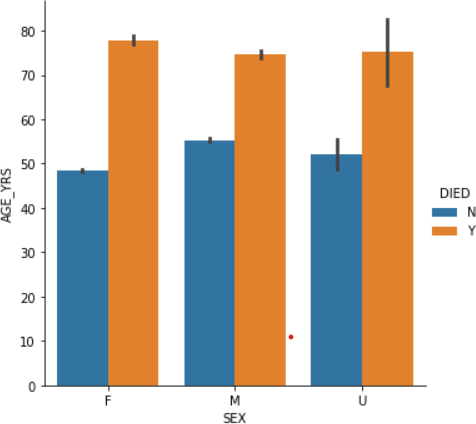
* Multicollinearity generally occurs when there are high correlations between two or more independent variables. In other words, one predictor variable can be used to predict the other. This creates redundant information, skewing the results in a regression model.
* We use the VIF factor for checking the multicollinearity. A variance inflation factor (VIF) detects multicollinearity in regression analysis. Multicollinearity is when there is a  correlation between predictors (i.e. independent variables) in a model; its presence can adversely affect your regression results.

## MULTIVARIATE ANALYSIS

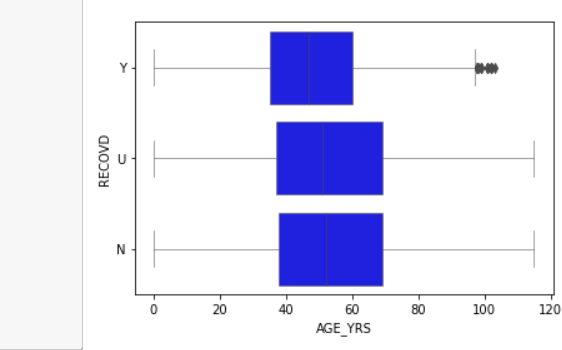
## [Multivariate Analysis](https://www.mygreatlearning.com/academy/learn-for-free/courses/multiple-variate-analysis?gl_blog_id=17681) is defined as a process of involving multiple dependent variables resulting in one outcome. This explains that the majority of the problems in the real world are Multivariate.



* The blue curve which shows the age of the people who have not DIED is almost normally distributed
* The green curve which shows the age of the people who have DIED is showing extreme values in the data (left skewed)



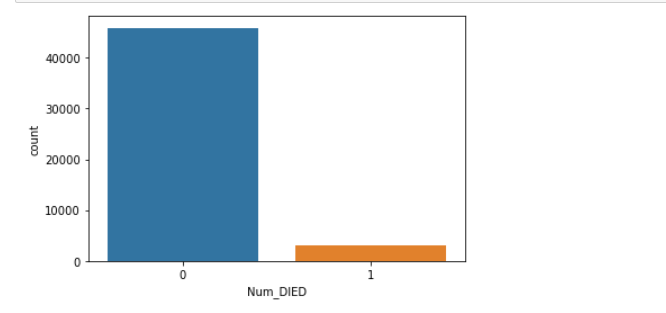
* + Female contains a large amount the datapoints. There is also a significant portion of unknown datapoints.
  + We can conclude that majority of the elderly female are dead.



* + We can see some extreme values in recovered data after age of 90

**FEATURE ENGINEERING**

**Class Imbalance and its Treatment:**

****

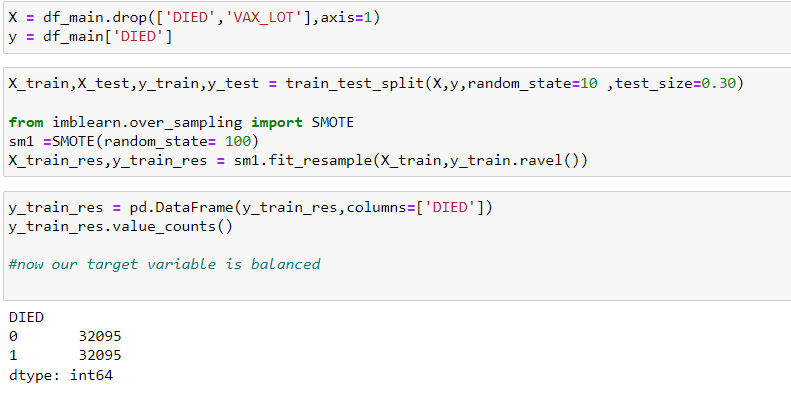
The target variable (Num\_died) is imbalanced

The challenge of working with imbalanced datasets is that most machine learning techniques will ignore, and in turn have poor performance on, the minority class, although typically it is performance on the minority class that is most important.

Here we have balanced the imbalanced target variable by using SMOTE Technique(

**Synthetic Minority Oversampling Technique**)

SMOTE first selects a minority class instance a at random and finds its k nearest minority class neighbors. The synthetic instance is then created by choosing one of the k nearest neighbors b at random and connecting a and b to form a line segment in the feature space. The synthetic instances are generated as a convex combination of the two chosen instances a and b



**Transformation for Categorical Variables:**

**One hot Encoding:** The target variable DIED column has performed One Hot Encoding and it is recorded in another column as Num\_Died.

**Weight of Evidence Encoding (WOE)**

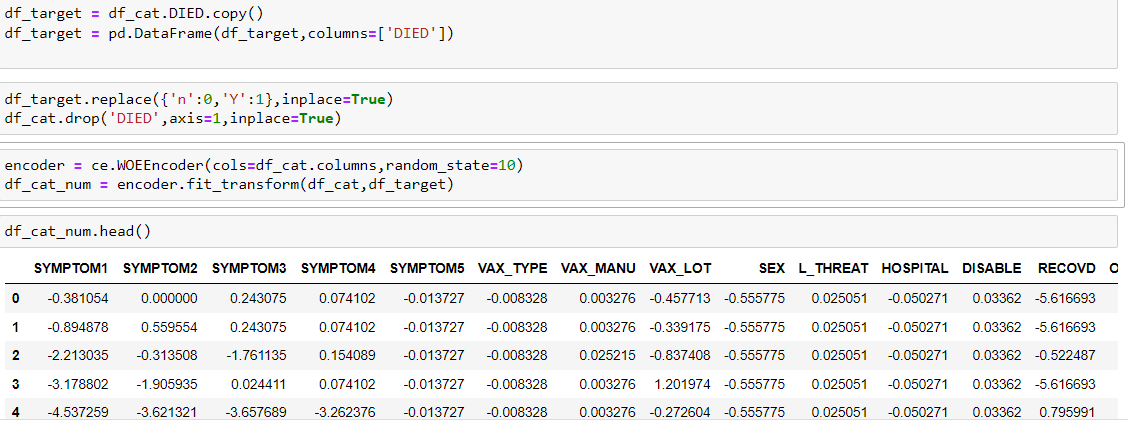
We have used woe encoding technique to encode the categorical variables

The weight of evidence tells the predictive power of an independent variable in relation to the dependent variable

We Combined categories with similar WOE and then create new categories of an independent variable with continuous WOE values. In other words, we used WOE values rather than raw categories in your model. The transformed variable will be a continuous variable with WOE values. It is same as any continuous variable.

**Rules related to WOE**

1. Each category (bin) should have at least 5% of the observations.
2. Each category (bin) should be non-zero for both non-events and events.
3. The WOE should be distinct for each category. Similar groups should be aggregated.
4. The WOE should be monotonic, that is either growing or decreasing with the groupings.
5. Missing values are binned separately.



**Data Preparation**

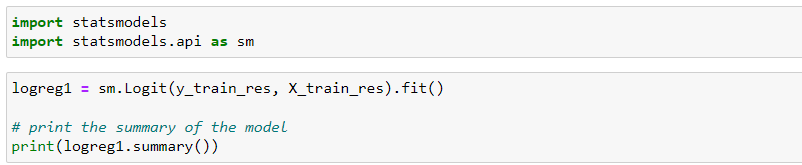
After preprocessing the train data, now the data is ready for modelling and evaluation. For this purpose, the train data and test data are split into two categories. Train Data is split into x\_train and ytrain and Test Data is split into x\_test and ytest.

**CHAPTER 5- CLASSIFICATION**

**5.1 Base Model building**

[Logistic Regression](https://en.wikipedia.org/wiki/Logistic_regression) is a Machine Learning classification algorithm that is used to predict the probability of a categorical dependent variable. In logistic regression, the dependent variable is a binary variable that contains data coded as 1 (yes, success, etc.) or 0 (no, failure, etc.). In other words, the logistic regression model predicts P(Y=1) as a function of X.

**Logistic Regression Assumptions:** dependent variable to be binary , the factor level 1 of the dependent variable should represent the desired outcome , only the meaningful variables should be included, that is, the model should have little or no multi-collinearity, the independent variables are linearly related to the log odds, logistic regression requires quite large sample sizes



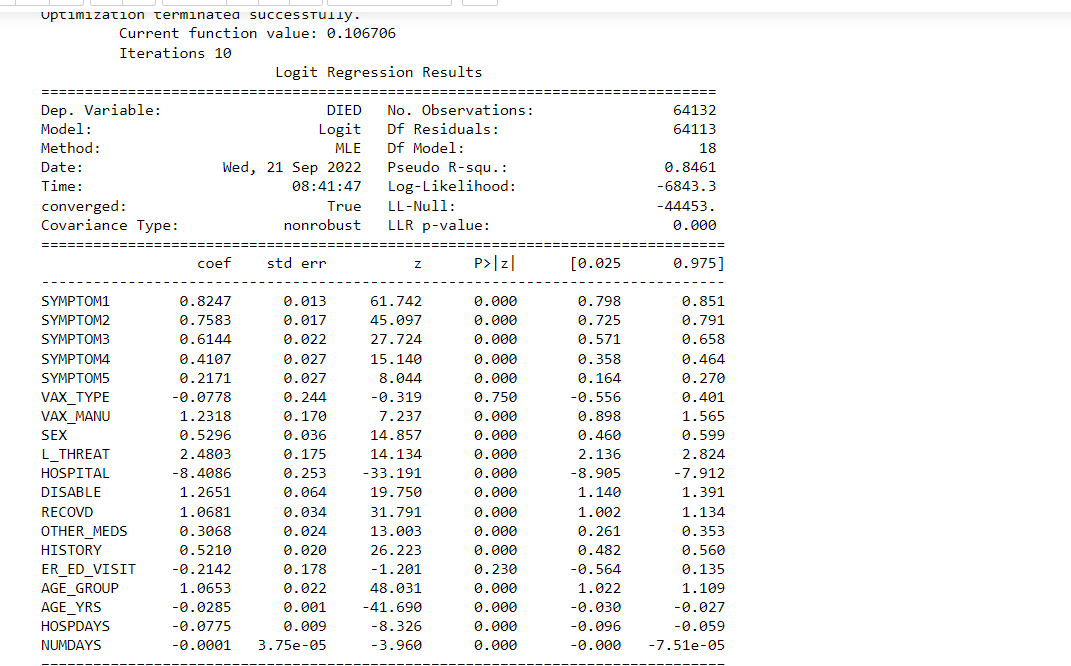


Fig.17:-Summary of Base Model

* **Log of Odds**

Log odds play an important role in logistic regression as it converts the LR model from probability based to a likelihood based model. Both probability and log odds have their own set of properties, however log odds makes interpreting the output easier. Thus, using log odds is slightly more advantageous over probability

1. Symptom1 = [2.281] it implies that the odds of detecting a threat of covid 19 vaccine increases by a factor of [2.281] due to one unit increase in the symptom1, keeping other variables constant
2. Symptom2 = [2.134] it implies that the odds of detecting a threat of covid 19 vaccine increases by a factor of [2.134] due to one unit increase in the symptom2, keeping other variables constant
3. Symptom3 = [1.848] it implies that the odds of detecting a threat of covid 19 vaccine increases by a factor of [1.848] due to one unit increase in the symptom3, keeping other variables constant
4. Symptom4 = [1.507] it implies that the odds of detecting a threat of covid 19 vaccine increases by a factor of [1.507] due to one unit increase in the symptom4, keeping other variables constant
5. Symptom5= [1.242] it implies that the odds of detecting a threat of covid 19 vaccine increases by a factor of [1.242] due to one unit increase in the symptom5, keeping other variables constant
6. VAX\_TYPE=[0.925] it implies that the odds of detecting a threat of covid 19 vaccine increases by a factor of [0.925] due to one unit increase in the VAX\_TYPE, keeping other variables constant
7. VAX\_MANU= [3.427] it implies that the odds of detecting a threat of covid 19 vaccine increases by a factor of [3.427] due to one unit increase in the VAX\_MANU, keeping other variables constant
8. SEX= [1.698] it implies that the odds of detecting a threat of covid 19 vaccine increases by a factor of [1.698] due to one unit increase in the SEX, keeping other variables constant
9. L\_THREAT= [11.944] it implies that the odds of detecting a threat of covid 19 vaccine increases by a factor of [11.944] due to one unit increase in the L\_THREAT keeping other variables constant
10. HOSPITAL= [0.0002] it implies that the odds of detecting a threat of covid 19 vaccine increases by a factor of [0.0002] due to one unit increase in the HOSPITAL , keeping other variables constant
11. DISABLE= [3.543] it implies that the odds of detecting a threat of covid 19 vaccine increases by a factor of [3.543] due to one unit increase in the DISABLE, keeping other variables constant
12. RECOVD= [2.909] it implies that the odds of detecting a threat of covid 19 vaccine increases by a factor of [2.909] due to one unit increase in the RECOVD, keeping other variables constant
13. OTHER\_MEDS= [1.359] it implies that the odds of detecting a threat of covid 19 vaccine increases by a factor of [1.359] due to one unit increase in the OTHER\_MEDS, keeping other variables constant
14. ER\_ED\_VISIT= [0.807] it implies that the odds of detecting a threat of covid 19 vaccine increases by a factor of [0.807] due to one unit increase in the ER\_ED\_VISIT, keeping other variables constant
15. HISTORY= [1.683] it implies that the odds of detecting a threat of covid 19 vaccine increases by a factor of [1.683] due to one unit increase in the HISTORY, keeping other variables constant
16. AGE\_GROUP= [2.901] it implies that the odds of detecting a threat of covid 19 vaccine increases by a factor of [2.901] due to one unit increase in the HISTORY, keeping other variables constant
17. AGE\_YRS= [0.971] it implies that the odds of detecting a threat of covid 19 vaccine increases by a factor of [0.971] due to one unit increase in the AGE\_YRS, keeping other variables constant
18. HOSPDAYS= [0.925] it implies that the odds of detecting a threat of covid 19 vaccine increases by a factor of [0.925] due to one unit increase in the HOSPDAYS, keeping other variables constant
19. NUMDAYS= [0.999] it implies that the odds of detecting a threat of covid 19 vaccine increases by a factor of [0.999] due to one unit increase in the NUMDAYS, keeping other variables constant

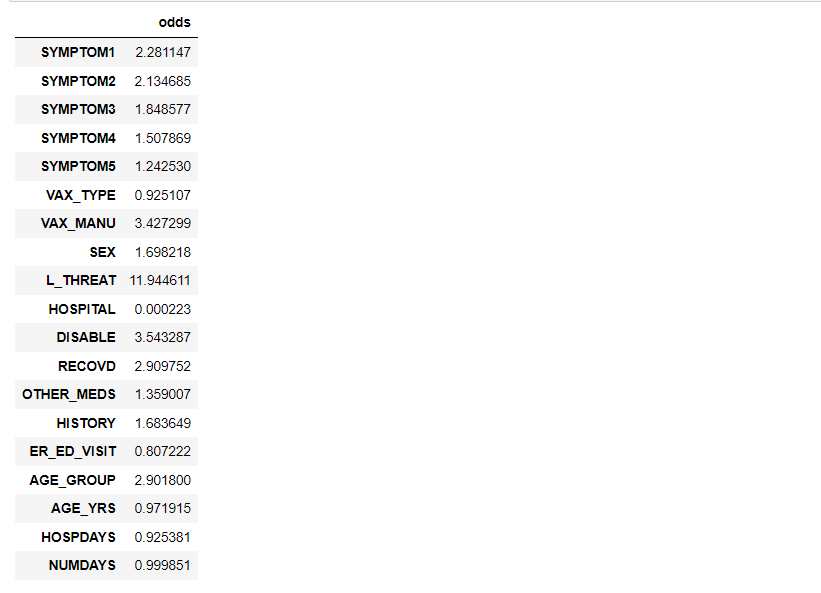
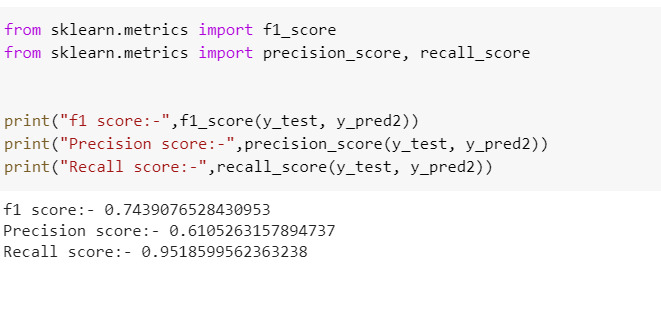


Fig.18:- Log of Odds

From Base Model we calculated the Performance Matrix and calculated the

the f1score,Precision Score,Recall Score



Inference:The f1 score for the base model is near about 1 which is a fairly good value but it can be improved.The precision score is 0.6 which is not a good value for a model but it also can be improved.Recall Score is 0.95 which is a good value for a model

Models we are going to perform:

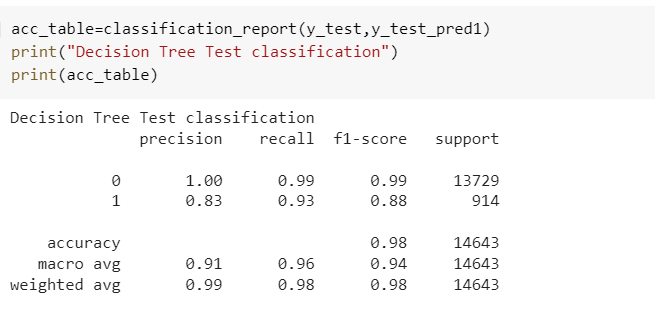
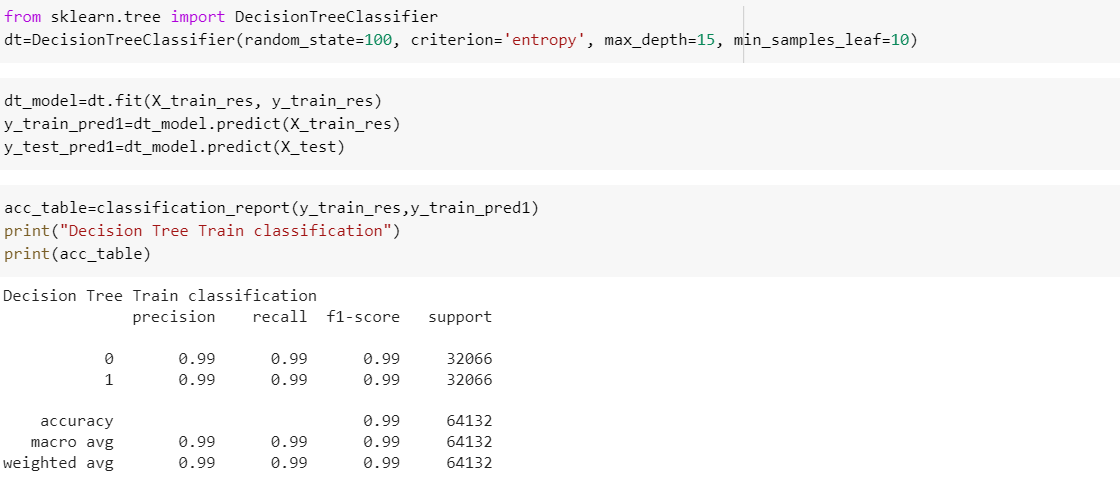
* Decision Tree Classifier
* Random Forest Classifier
* Ada Boosting
* Gradient Boosting

**CHAPTER 6-MODEL BUILDING**

* 1. **DECISION TREE**

The first model used for our project is Decision Tree Classifier. The reason for selecting Decision Tree is that it is not sensitive to outliers since outliers never cause much reduction in Residual Sum of Squares (RSS) because they are never involved in the split. In our project we did not remove or cap the outliers because every data point is important for prediction and cannot treat the outliers until the client allows us to do the same. There is no requirement of feature scaling techniques such as standardization and normalization as it uses a rule-based approach instead of calculation of distances. Decision Trees are a class of very powerful Machine Learning models cable of achieving high accuracy in many tasks while being highly interpretable. What makes decision trees special in the realm of ML models is really their clarity of information representation.

* **IMPORTING THE LIBRARIES AND SETTING THE PARAMETERS**

****

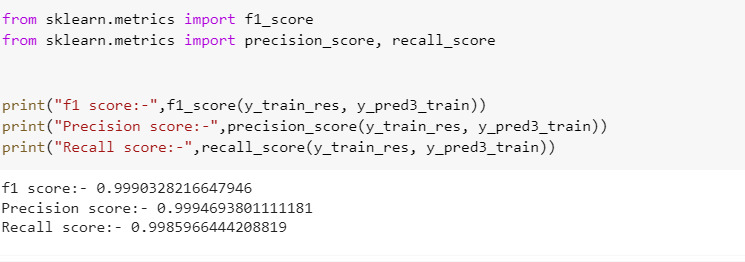
Inference:f1 score gives the harmonic mean of Precision and Recall.Higher the f1 score better will be the model(0 to 1).so here the f1 score is 0.99 for train set and 0.98 for test set which is a good for the model

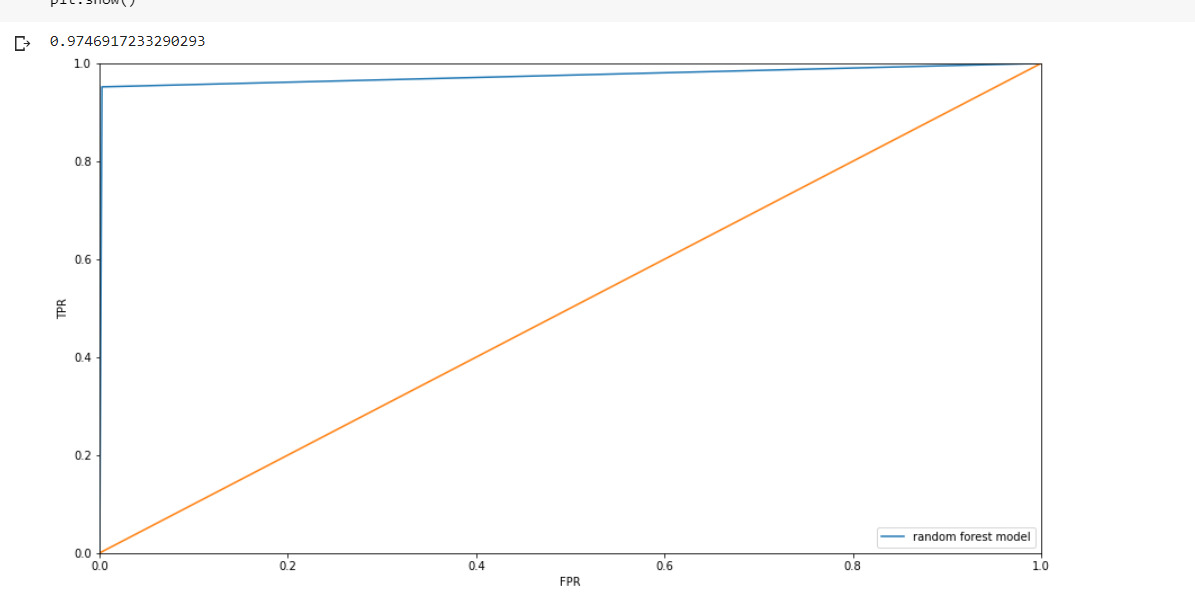
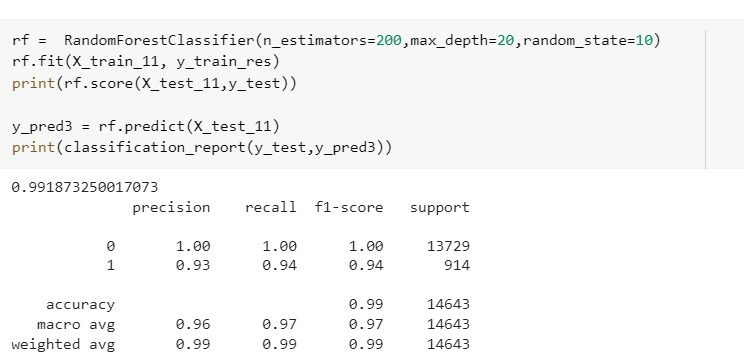
* 1. **RANDOM FOREST CLASSIFIER**

It is a supervised machine learning algorithm that is widely used in classification problems that consists of many decision trees. It uses bagging and feature randomness when building each individual tree to try to create an uncorrelated forest of trees whose prediction by committee is more accurate than any individual tree.

Random forest generally reduces the bias error by allocating the features and reduce the variance error by bootstrap sampling or bagging technique. This bootstrapping procedure leads to better model performance because it decreases the variance of the model, without increasing the bias. .

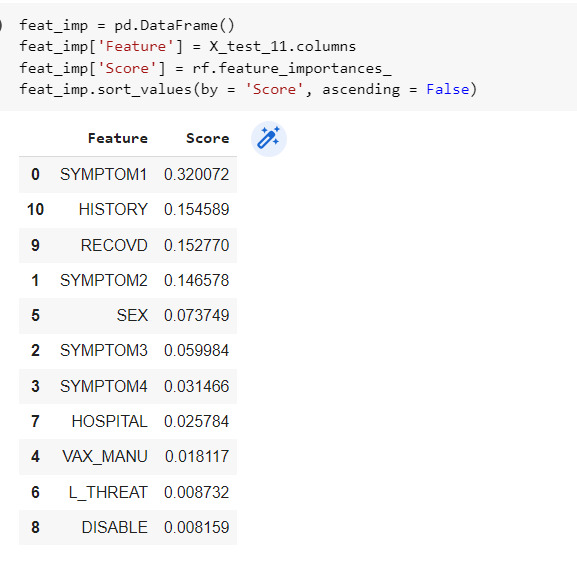
* **IMPORTING THE LIBRARIES AND SETTING THE PARAMETERS**

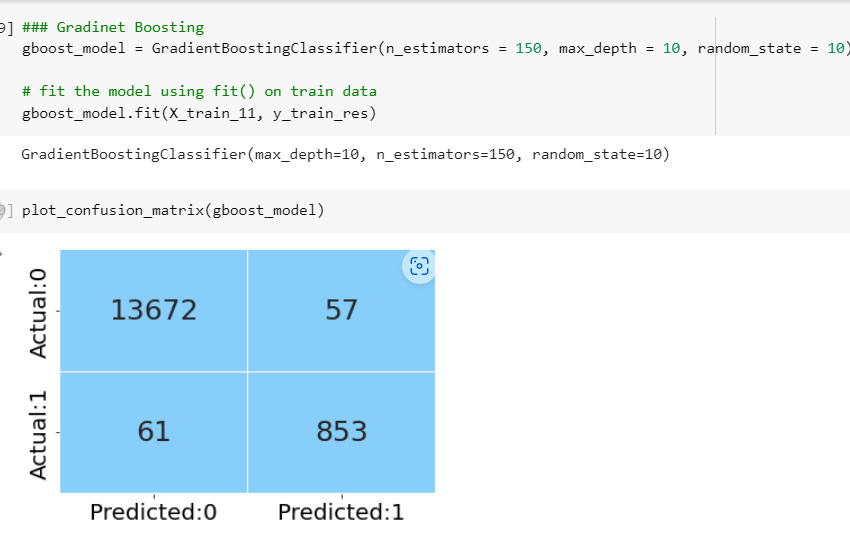
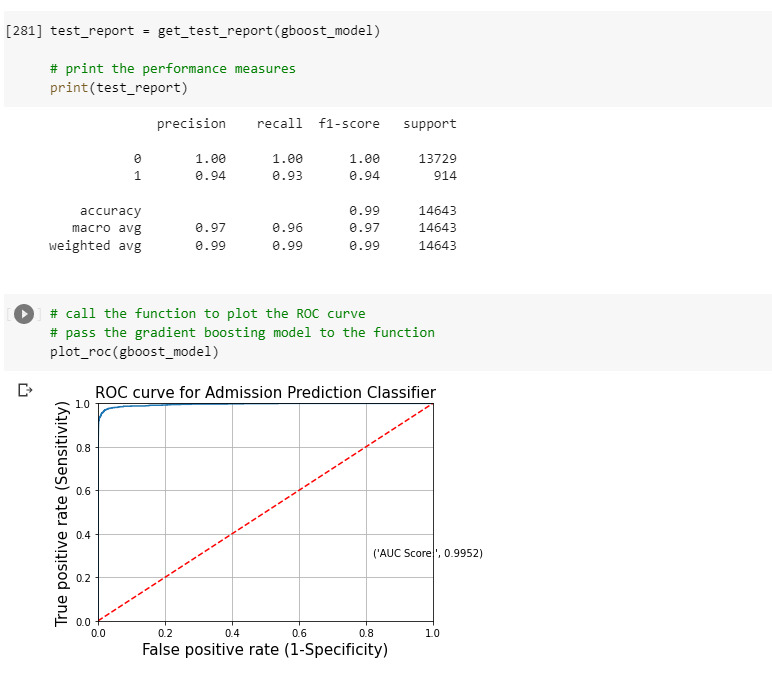
****

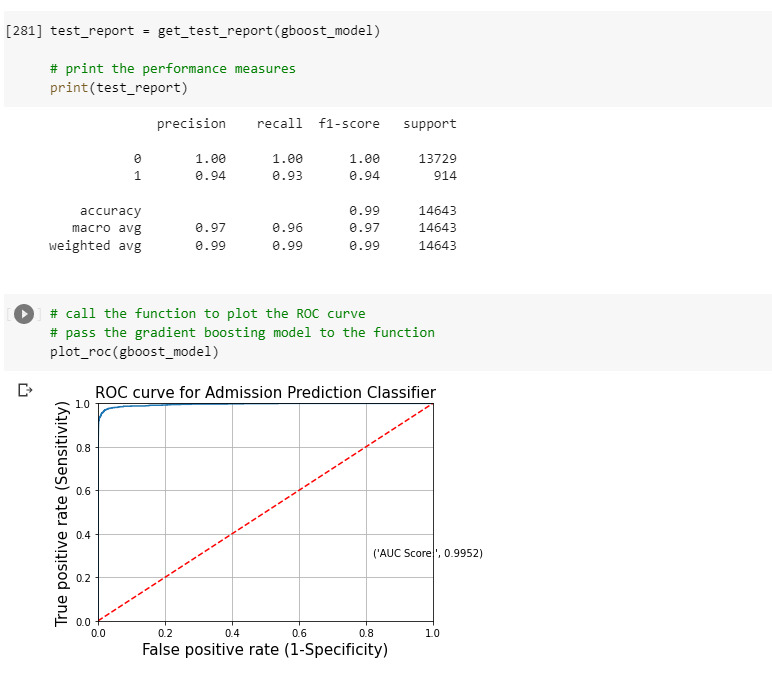


Roc score is 0.97

* Important features after doing Random Forest







Inference:

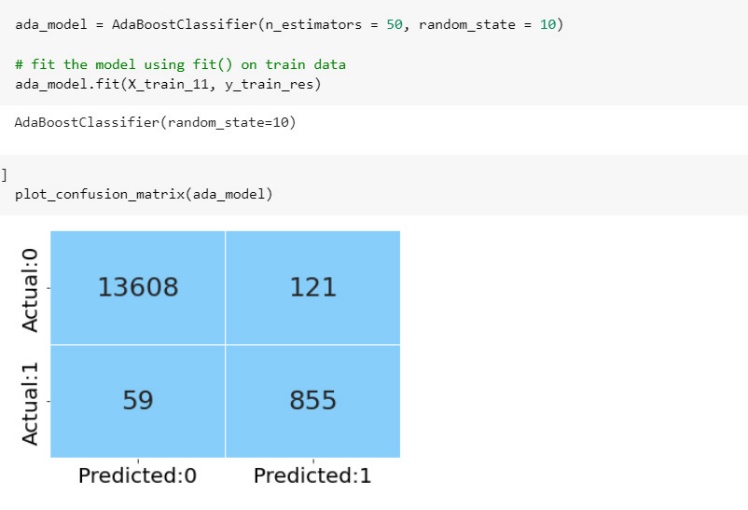
The f1 score for the model after done the gradient boosting is 0.99

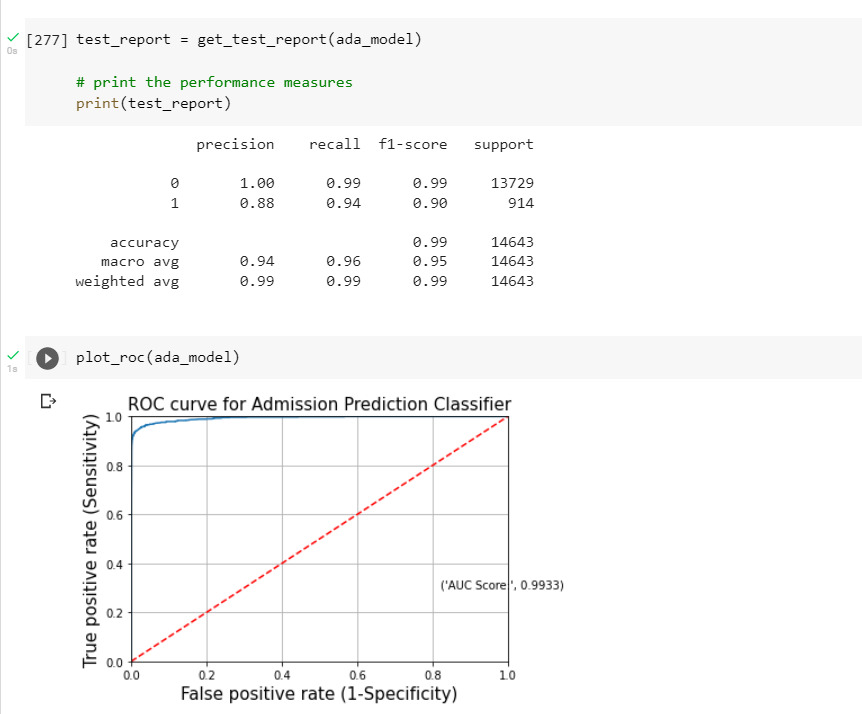
Which is a good value for the model.

**ROC CURVE WITH AUC SCORE**

The AUC score for the test data is 0.9969. When the AUC is greater than 0.5, there is a high chance that the classifier will be able to distinguish the positive and negative classes and further we should try to maximize the AUC score.

Ada Boosting





|  |  |  |  |
| --- | --- | --- | --- |
| Model | Precision | F1-score | Recall |
| Decision Tree(Train) | 0.99 | 0.99 | 0.99 |
| Decision Tree(Test) | 0.91 | 0.94 | 0.96 |
| Random Forest(Train) | 0.99 | 0.99 | 0.99 |
| Random Forest(Test) | 0.96 | 0.95 | 0.95 |
| Gradient Boosting(Test) | 0.97 | 0.97 | 0.96 |
| Ada Boosting (Test) | 0.94 | 0.95 | 0.96 |
|  |  |  |  |
|  |  |  |  |

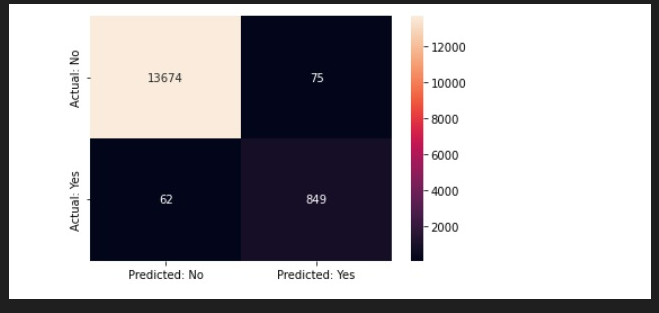
**How does your solution affect the problem in the domain or business? What**

**recommendations would you make, and with what level of confidence?**

**What are the limitations of your solution? Where does your model fall short in**

**the real world? What can you do to enhance the solution?**

* The models which we have built till now is giving good accuracy score, roc\_auc score, F1 score .so this real time dataset can be used for further predictions.



* We can do feature selection.
* We can do hyper parameter tuning to enhance our model.



CONCLUSION

In the healthcare sector, a huge amount of data is generated continuously and this data can be used to extract meaningful information. The main objective of our project is to work towards building a suitable model to predict whether a patient’s life will be endangered after taking the Covid-19 vaccine.



REFERENCES

The references can be blogs, articles or even social media news relevant to explain the importance of the projects.

**https://**[**www.kaggle.com/datasets/ayushggarg/covid19-vaccine-adverse-**](http://www.kaggle.com/datasets/ayushggarg/covid19-vaccine-adverse-) **reactions?resource=download&select=2021VAERSDATA.csv**

**Conclusion:-**

**In the healthcare sector, a huge amount of data is generated continuously and this data can be used to extract**

**meaningful information. Through our project we have tried to predict if someone is going to die from the side effect of the**

**Covid-19 Vaccination. We have made a good quality model for this purpous which can predict this.**

**Although we have some wrong prediction while testing but that is well under the limit and most of the prediction were correct.**

**So through our model we can actually predict if someone's life is at resk due to the ill effect of vaccination and we can**

**start to treat him or her accordingly before hand to save their lives.**