SC1015

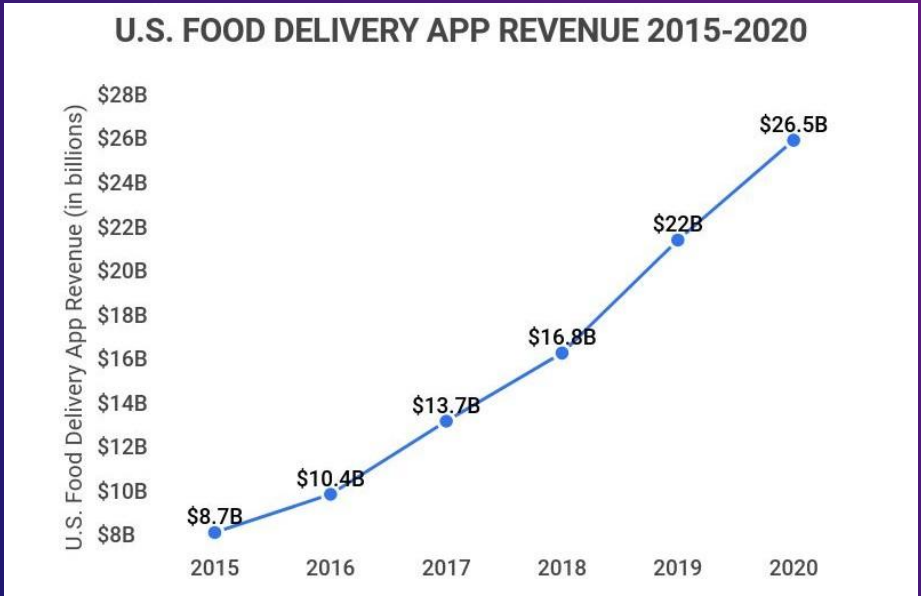
MINI PROJECT

ONLINE FOOD ORDER DATASET

LAB GROUP FCSI, TEAM 6

KOH JIA XIN (U2322256D)
MATHEW ALAN SHILU (U2323425E)
GIRISH PRANAV (U2323962J)

ONLINE FOOD DELIVERY MARKET IN U.S.



AS OF 2020, THE ONLINE FOOD DELIVERY APP INDUSTRY MADE $26.5 BILLION (USD) IN REVENUE

AS OF 2014, ONLINE FOOD ORDERING GROWN 300% FASTER THAN DINE-IN

PROBLEM DEFINITION

HOW DIFFERENT VARIABLES AFFECT THE FEEDBACK GIVEN FOR THE ONLINE FOOD ORDER?

01

WHY DOES FEEDBACK MATTER?

Feedback provides valuable insights into consumer preferences and influences purchasing decisions, thereby impacting brand reputation and customer retention. Analysing feedback helps businesses identify areas for improvement and tailor their services to enhance the overall customer experience, ensuring competitiveness in the market.

02

WHAT ARE THE VARIABLES?

1. Age
2. Family size
3. Occupation
4. Marital Status
5. Monthly Income

EXPLORATORY DATA ANALYSIS

AGE, MONTHLY INCOME, OCCUPATION, MARITAL STATUS, FAMILY SIZE

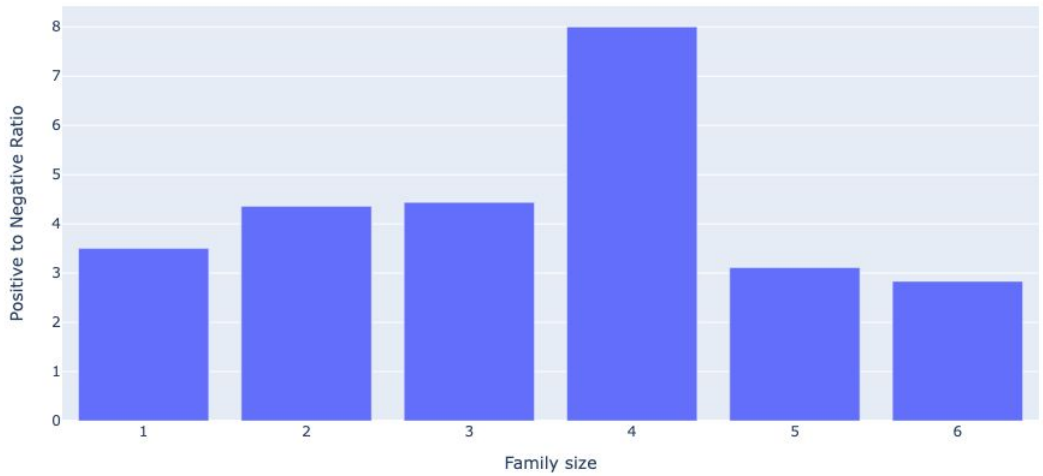
AGE

YOUNGER CUSTOMERS, TEND TO PROVIDE MORE POSITIVE FEEDBACK, WHILE NEGATIVE FEEDBACK TENDS TO COME FROM SLIGHTLY OLDER CUSTOMERS

FAMILY SIZE

IT SEEMS LIKE THE FAMILY OF 3 GIVES THE HIGHEST NUMBER OF POSITIVE OUTCOMES. HOWEVER, IT ALSO HAS THE HIGHEST NUMBER OF NEGATIVE OUTCOMES. WE SHOULD EXAMINE THE POSITIVE TO NEGATIVE RATIO FOR EACH FAMILY SIZE.

FAMILY SIZE

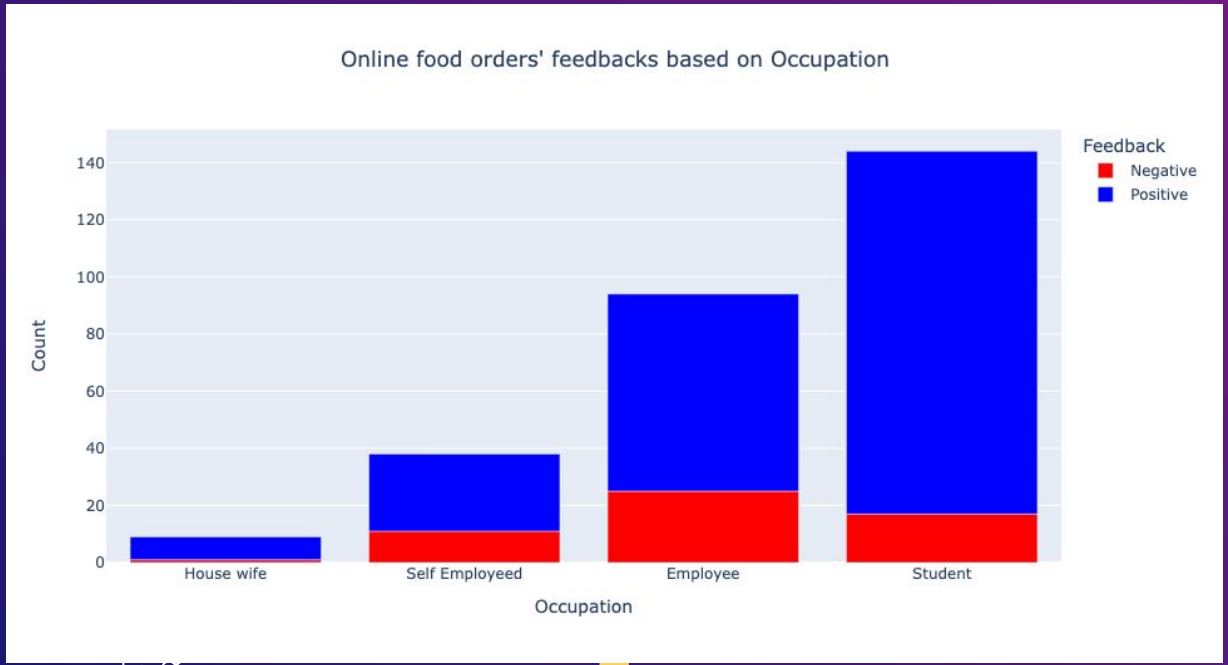


FAMILY OF 4 SHOW A HIGHER POSITIVE TO NEGATIVE FEEDBACK RATIO. POSSIBLE REASONS MAY BE COST EFFECTIVENESS AND MEAL SHARING, LEADING TO MORE POSITIVE FEEDBACK GIVEN

MONTHLY INCOME

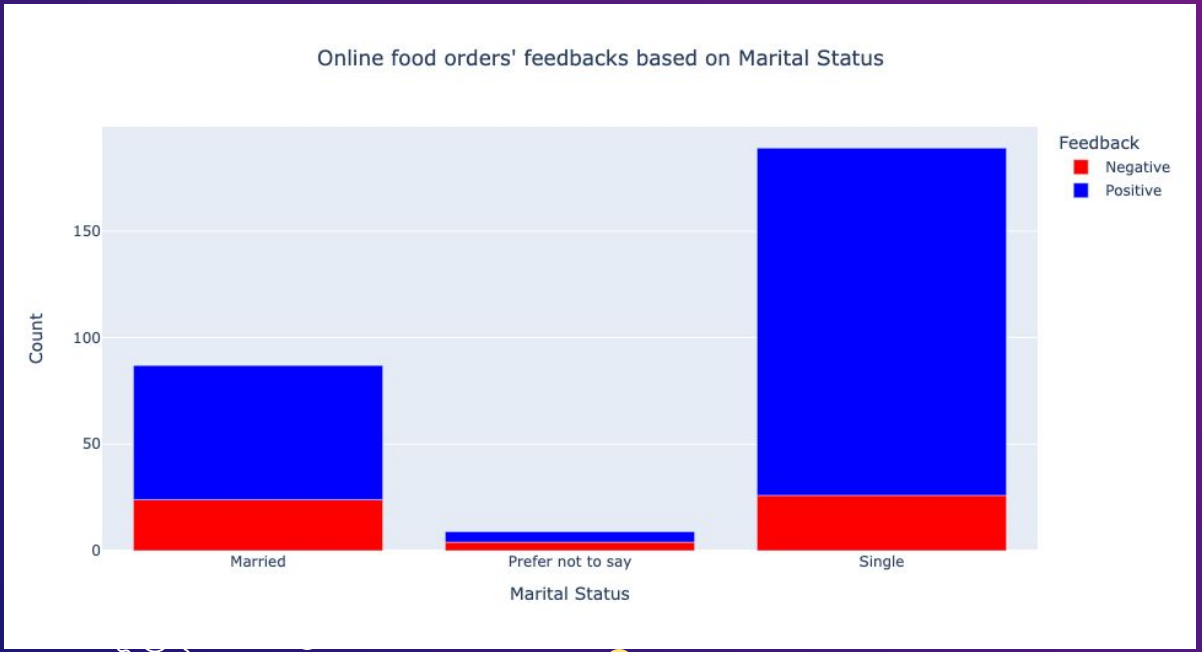
THERE IS A SLIGHT INCREMENT IN POSITIVE FEEDBACK AS MONTHLY INCOME INCREASES. HOWEVER, THE CATEGORY "NO INCOME" GIVES THE MOST POSITIVE FEEDBACK

OCCUPATION



STUDENTS GIVES MORE POSITIVE FEEDBACK. HOWEVER, IT COULD BE INFLUENCED BY OTHER FACTORS SUCH AS BUDGET CONSTRAINTS, CONVENIENCE AND LIMITED ALTERNATIVES IN SCHOOL

MARITAL STATUS



SINGLES PROVIDE MORE POSITIVE FEEDBACK COMPARED TO MARRIED. MARITAL STATUS MAY HAVE AN INFLUENCE ON THE LIKELIHOOD OF LEAVING POSITIVE FEEDBACK FOR ONLINE FOOD ORDERS

PROBLEM DEFINITION

HOW DIFFERENT VARIABLES SUCH AS AGE, FAMILY SIZE, MONTHLY INCOME, OCCUPATION AND MARITAL STATUS AFFECT THE FEEDBACK GIVEN FOR THE ONLINE FOOD ORDER?

DATA PREPARATION AND CLEANING

01 Remove duplicates, unnamed column and insignificant columns

02 Ensure that there is no null values

03 Generate descriptive statistics for numerical columns

04 Encoding categorical variables

05 Upsampling *(which will be explained later)*

REMOVE DUPLICATES, UNNAMED AND INSIGNIFICANT COLUMNS

	Age	Gender	Marital Status	Occupation	Monthly Income	Educational Qualifications	Family size	latitude	longitude	Pin code	Output	Feedback	Unnamed: 12
0	20	Female	Single	Student	No Income	Post Graduate	4	12.9766	77.5993	560001	Yes	Positive	Yes
1	24	Female	Single	Student	Below Rs.10000	Graduate	3	12.9770	77.5773	560009	Yes	Positive	Yes
2	22	Male	Single	Student	Below Rs.10000	Post Graduate	3	12.9551	77.6593	560017	Yes	Negative	Yes
3	22	Female	Single	Student	No Income	Graduate	6	12.9473	77.5616	560019	Yes	Positive	Yes
4	22	Male	Single	Student	Below Rs.10000	Post Graduate	4	12.9850	77.5533	560010	Yes	Positive	Yes

	Age	Gender	Marital Status	Occupation	Monthly Income	Educational Qualifications	Family size	Output	Feedback
0	20	Female	Single	Student	No Income	Post Graduate	4	Yes	Positive
1	24	Female	Single	Student	Below Rs.10000	Graduate	3	Yes	Positive
2	22	Male	Single	Student	Below Rs.10000	Post Graduate	3	Yes	Negative
3	22	Female	Single	Student	No Income	Graduate	6	Yes	Positive
4	22	Male	Single	Student	Below Rs.10000	Post Graduate	4	Yes	Positive

ENCODING CATEGORICAL VARIABLE

MARITAL STATUS

```
{'Married': 0, 'Prefer not to say': 1, 'Single': 2}
  Marital Status  Marital Status Encoded
0         Single                       2
1         Single                       2
2         Single                       2
3         Single                       2
4         Single                       2
```

OCCUPATION

```
{'Employee': 0, 'House wife': 1, 'Self Employeed': 2, 'Student': 3}
  Occupation  Occupation Encoded
0    Student                   3
1    Student                   3
2    Student                   3
3    Student                   3
4    Student                   3
```

MONTHLY INCOME

```
{'No Income': 0, 'Below Rs.10000': 1, '10001 to 25000': 2, '25001 to 50000': 3, 'More than 50000': 4}
   Monthly Income Monthly Income Encoded
0       No Income                      0
1  Below Rs.10000                      1
2  Below Rs.10000                      1
3       No Income                      0
4  Below Rs.10000                      1
```

FEEDBACK

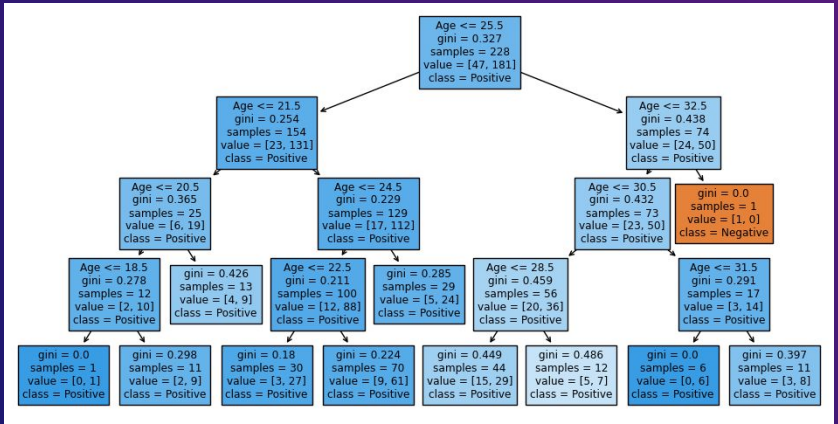
```
{'Negative ': 0, 'Positive': 1}
    Feedback  Feedback Encoded
0   Positive                 1
1   Positive                 1
2  Negative                  0
3   Positive                 1
4   Positive                 1
```

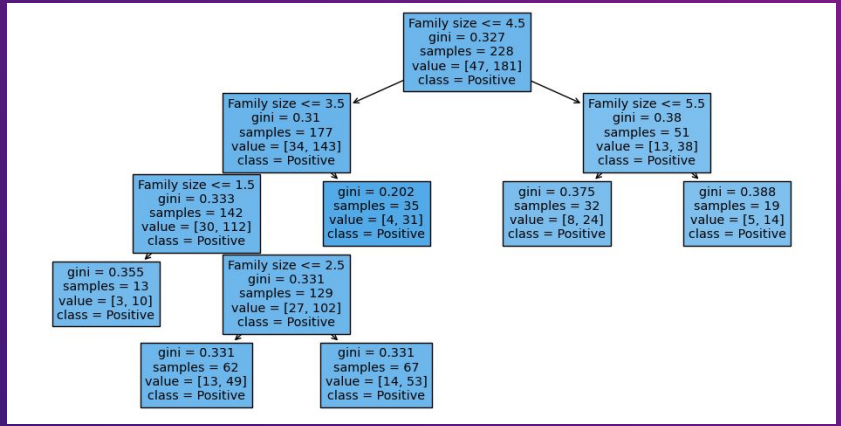
MACHINE LEARNING

- DECISION TREE
- RANDOM FOREST
- LOGISTICS REGRESSION

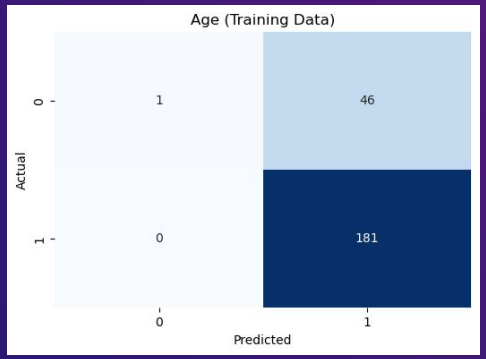
DECISION TREE

- CLASSIFY CATEGORICAL RESPONSE (FEEDBACK) USING NUMERICAL PREDICTORS (AGE AND FAMILY SIZE)
- ACCURACY IS ABOVE 79% AND TRUE POSITIVE RATE IS HIGH AT 1.0
- FALSE POSITIVE RATE IS HIGH, ABOVE 79%





BEFORE UPSAMPLING

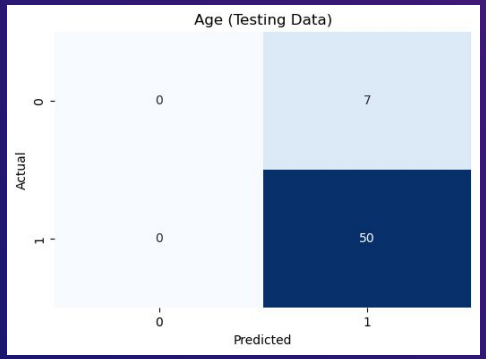


TRAIN DATA

ACCURACY: 0.798

TRUE POSITIVE RATE: 1.0

FALSE POSITIVE RATE: 0.978



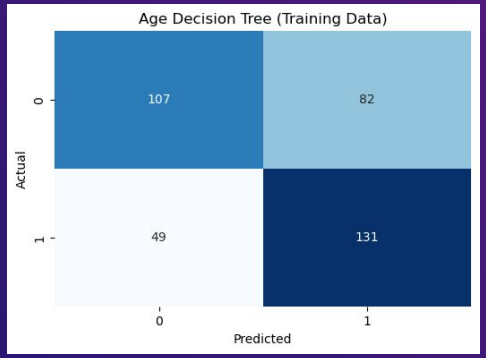
TEST DATA

ACCURACY: 0.877

TRUE POSITIVE RATE: 1.0

FALSE POSITIVE RATE: 1.0

AFTER UPSAMPLING

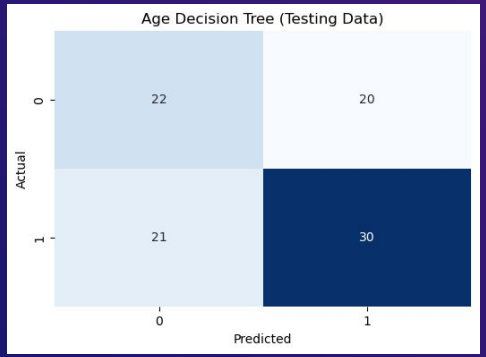


TRAIN DATA

ACCURACY: 0.542

TRUE POSITIVE RATE: 0.555

FALSE POSITIVE RATE: 0.470



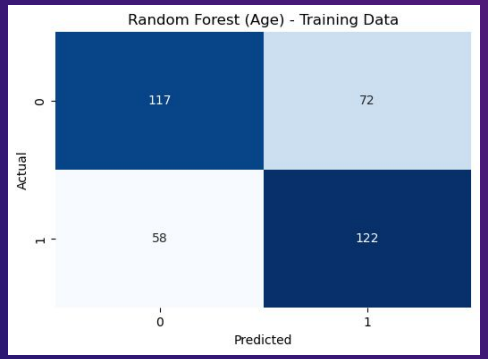
TEST DATA

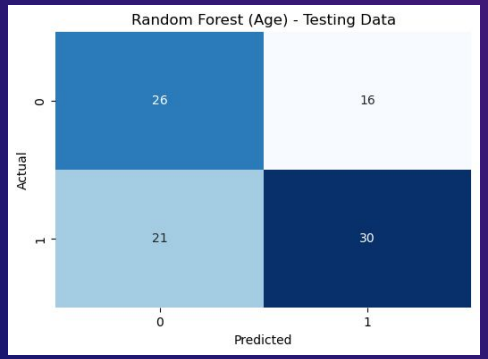
ACCURACY: 0.603

TRUE POSITIVE RATE: 0.568

FALSE POSITIVE RATE: 0.357

RANDOM FOREST





TRAIN DATA

ACCURACY: 0.648

TRUE POSITIVE RATE: 0.678

FALSE POSITIVE RATE: 0.381

TEST DATA

ACCURACY: 0.602

TRUE POSITIVE RATE: 0.588

FALSE POSITIVE RATE: 0.381

- MODERATE ACCURACY AND TRUE POSITIVE RATE
- REDUCTION IN THE DISCREPANCY BETWEEN THE TRAIN AND TEST DATASETS
- IMPROVED PERFORMANCE ACROSS THE DATA

LOGISTIC REGRESSION

- THE MODEL ACHIEVED AN ACCURACY OF 80% ON TRAIN SET AND 86% ON TEST SET

- TOP THREE CATEGORICAL PREDICTORS INFLUENCING POSITIVE FEEDBACK OUTCOMES:

1. OCCUPATION: HOUSEWIFE
2. MONTHLY INCOME: MORE THAN 50000
3. MARITAL STATUS: SINGLE

WHAT HAVE WE LEARNT

1. USING A NEW MACHINE LEARNING FUNCTION: RANDOM FOREST
2. UNDERSTANDING OF REAL WORLD APPLICATION OF MACHINE LEARNING IN ADDRESSING BUSINESS PROBLEMS

OUTCOME OF OUR PROJECT

1. ONLINE FOOD ORDERING COMPANIES CAN USE OUR MODEL TO PREDICT THEIR FEEDBACK
2. PREDICT WHAT CHANGES CAN BE MADE TO IMPROVE THEIR FEEDBACK
3. IMPROVE CUSTOMER SATISFACTION

CONCLUSION AND INSIGHTS

FAMILIES WITH MORE THAN 4 MEMBERS (5 TO 6) TEND TO GIVE FEWER POSITIVE FEEDBACK

- OFFER VALUE PACKED BUNDLE MEALS FOR LARGER FAMILIES

SINGLES TEND TO GIVE MORE POSITIVE FEEDBACK THAN MARRIED INDIVIDUALS

- INTRODUCE FAMILY-FRIENDLY MENU OPTIONS OR MEAL BUNDLES THAT CATER TO THE DIETARY PREFERENCES OF MARRIED COUPLES

EMPLOYEE AND SELF-EMPLOYED HAS FEWER TENDENCY TO GIVE POSITIVE FEEDBACK

- STREAMLINE THE ORDERING AND DELIVERY PROCESS TO CATER TO THE BUSY SCHEDULE OF EMPLOYED INDIVIDUALS

REFERENCES

Dharmavarapu H. R. S. S. S. S. Manikanth. (2020). Impact of the Online Customer Reviews in Consumer's Food Ordering Decision for the Online Food Delivery Service Apps. *International Journal of Advanced Science and Technology*, *29*(5s), 2692-2703. Retrieved from http://sersc.org/journals/index.php/IJAST/article/view/22864

Zippia. 18+ Food Delivery Statistics (2023). Online Ordering Industry Numbers You Need To Know. Retrieved from https://www.zippia.com/advice/food-delivery-industry-statistics/

Sruthi, E. R. (2024). Understand Random Forest Algorithms With Examples. Retrieved from https://www.analyticsvidhya.com/blog/2021/06/understanding-random-forest/

Kanade, V. (2022). What Is Logistic Regression? Equation, Assumptions, Types, and Best Practices. Retrieved from https://www.spiceworks.com/tech/artificial-intelligence/articles/what-is-logistic-regression/