

The Economic Impact of Coronavirus in Sub-Saharan Africa

Swaranika Palit (UW ID : 20827180)

Sri Priyanka Kommalapati (UW ID : 20841127)

Department of Management Sciences, University of Waterloo

MSCI 623 : Big Data Analytics

August 5, 2020

ABSTRACT

This project analyzes the economic impact of COVID-19 in Sub-Saharan Africa by predicting how concerned are the people in Africa about paying their expenses for the next six months due to COVID-19, based on different attributes like age, gender, monthly need, monthly income, employment type. Also, job loss due to COVID-19, the kind of aid people are receiving or any loan they are currently paying now or any new loan they have taken to pay the expenses due to COVID-19, if the person is responsible for the household responsibility, mobile money activity are also considered in the estimation of the economic expense concern in Africa. Additionally, what people are thinking about the priority action for the Government are also important in estimating the expense concern of people of Africa. Also, we identify some patterns and rules in the data that can help the government in prioritizing their upcoming decisions.

The data mining analysis has been done on the dataset provided as part of the survey of Economic Impact of COVID-19 in Sub-Saharan Africa conducted by **GeoPoll** and published on the HUMANITARIAN DATA EXCHANGE(HDX) website. As a part of the supervised data-mining process, we have used different classifiers model to predict the rating provided by the person in terms of 1 to 5 for their expense concern rating (starting from no concerned at all to extremely concerned). From the model output result, we have noticed that majority of the people have expressed their concern with the highest rating due to COVID-19 and there are few people who have expressed concern in medium range. This finding addresses the issue that majority of people in Sub-Saharan Africa are affected due to the pandemic and are highly concerned about the expenses for next six months. We have also noticed from our classifier models that age, monthly income, concern of expenses of the current month, monthly need, if the person concerned is responsible for the household expenses, expected duration of the time in terms of the ability to pay for basic expenses like rent or food, aid received, the source of money for expenses are in in form of loan or credits and if the person is able to deposit mobile money – these attributes are most important in predicting the concern expense rating for the upcoming six months.

From the unsupervised learning using association, we have found that few of the frequent item sets are: Aid not received, Has Expense Responsibility of the household, Job lost due to COVID-19, More concerned about expenses than before COVID-19. Also, few of the rules with high confidence generated are {Aid Not Received <-> Has Expense Responsibility}, {Aid Not Received <-> Job lost due to Covid-19}. Most of the generated rules from the dataset are the combination of above rules. Although the rules generated have no strong association, we have found some important patterns in the data. Hence, we can say from the rule findings that most employed people have lost their jobs due to COVID-19 and these people have expense responsibility of their household and had not received any aid from any organization.

INTRODUCTION

As COVID-19 continues to spread around the world leaving economic upheaval in its wake, we have focused on analyzing the economic impact of Sub-Saharan Africa which will eventually help the concerned government to build a road plan. Specifically, government can introduce more aids and instantiate laid off workers from traditional job markets to a post-pandemic job markets specially for those people who are highly concerned about carrying out their expenses for the upcoming months.

Here, as we are using supervised learning to classify people of Sub-Saharan Africa in terms of their expression of expense concern due to COVID-19 for the upcoming months based on the social factors, the most important features which have been used for predicting the concern expense rating will also provide important insights for the government while taking consideration of any new initiative to address the needy people and build the economy.

We use association rule mining of unsupervised learning to identify the most frequent attributes and the common rules observed in the dataset. These frequent features can help to target the people who should be the primary focus from the concerned Government's program in the process of rebuilding the economy in post COVID-19 time. Our findings from the classifier model and the association rule mining can contribute to help the government of local province in Sub-Saharan Africa as well as country wise to construct their economic rebuilding roadmaps in the post pandemic time and also provide them important insights to be used for their different aid programs as well targeting the correct people.

We believe that post COVID-19, it is very important to address specific people who are in mostly need of economic help to instantiate the economy and this data-mining project will be useful in that aspect. Finding a solution to the ongoing crisis makes this project interesting. Developing a model to find the economic concern of a person based on social factors of the person and using rule mining to find out the association between these factors makes our project novel.

In the following section, we will be going to perform two data-mining tasks: (a) for supervised learning , we will build the classifier model to estimate the rating which is expressed on the scale of 1 to 5 for the concern of expenses in upcoming six months and (b) unsupervised learning – we will use association rule mining on different combination of the available features (with different support and confidence values) and identify association rules of various socio-economic factors. Here we have not used any clustering algorithm as most of our features in the dataset are categorical and we feel that clustering will not be that meaningful to interpret some useful insights from the data.

For both the algorithms we start by considering all features and proceed by eliminating the variables that may not be important to us. The detailed description of type of variables used is mentioned in the corresponding algorithm implementation sections.

RELATED WORK

We have reviewed few research papers on the economic impact due to COVID-19 not only on Africa but also on other countries. Most of the papers address the economic impact either on different business sectors or use GDP as part of the estimations. Our dataset is primarily focused on individual survey analyzing their concern for the expenses, predicting the impact of economy on individual level due to COVID-19.

The previous work report^[1] which has been conducted based on the same dataset has mostly focused on the overall attributes like their percentage value in each label of those features with visualization. Though this report will provide the overview of the features, this does not include any work on estimating how much concerned people are in terms managing their expenses based on the other available features. Apart from the prediction, our data mining project using this dataset focuses on finding the important attributes and mining insightful rules which have not been covered earlier. By finding these association rules, we can interpret the features in more meaningful way and most importantly these association of features can attribute to the Government initiative of recovering the economy post pandemic by streamlining the focused people.

The paper “COVID 19 Pandemic, a War to be Won: Understanding its Economic Implications for Africa” by John E. Ataguba^[2] has primarily focused on the microeconomic costs and demonstrated that any restriction or removal of the ability to work and earn a living, will put economic pressure on family. It also highlights that when export-oriented firms or enterprises are unable to export goods, the demand for exports and export income will decline, which is accompanied by the downsizing of production units and the laying of workers.

The paper “COVID-19 in Africa: Protecting Lives and Economies” by the Economic Commission for Africa^[3] focuses on the four type of challenges Africa is encountering in terms of the economic impact due to the pandemic - high debt-to-GDP levels, high fiscal deficits; high costs of borrowing and depreciation of many African currencies against the euro and the dollar.

The paper, “COVID induced economic uncertainty” by National Bureau of Economic Research^[4] has also focused on three indicators – stock market volatility, newspaper-based economic uncertainty, and subjective uncertainty in business expectation surveys and use these indicators to document and quantify the enormous increase in economic uncertainty during the pandemic.

In the case study “Estimating the Economic Impact of COVID-19: A Case Study of Namibia” by National Planning Commission, University of Namibia^[5] has narrated the impact of COVID-19 economic impact based on different sectors and proposed a theoretical framework as $GDP = C + I + G + (X - M)$ where, C is household consumption of domestically produced goods and services, I is investment purchases by firms of domestically produced goods and services, G is government purchases of domestically produced goods and services, $X - M$ is exports(X) minus

imports (M), where the difference is the trade balance or the net export. The outbreak COVID-19 by implication have impacts on consumption because of lost lives, employment, and income.

In the paper “Understanding Coronanomics: The economic implications of the coronavirus (COVID-19) pandemic” by Suborna Barua^[6], we have observed a comprehensive and indicative overview on the observed and the possible economic impacts – a general mapping of the likely economic impacts of COVID-19 has been introduced and also no quantitative estimate of the future impacts has not been produced .

The paper “COVID-19 Outbreak: Is it a Health Crisis or Economic Crisis or Both? Case of African Counties” by Jacob Joseph Kassema^[7] provides an explorative research design with reviewed data from the online survey and various articles and journals about the impact of COVID-19 outbreak to African developing countries in the area of human capital, gross domestic product, social life, productivity as well as business. It has been projected that the crisis will cut off 1.4% of the USD 2.1 Trillion cumulative GDP of Africa due to an encompassing destruction of businesses on the continent and across the globe.

The paper ‘Impact of Covid-19 on the South African economy’ published by the government of South Africa^[8] in collaboration between local and international research institutes and the government of South Africa has used Social Accounting Matrix (SAM) to link the corresponding income and expenditures of all economic actors for exploring the economy-wide effects of shocks to the economy, each with their own strengths and limitations.

In the paper –‘Real-Time Economics: A New Platform to Track the Impacts of COVID-19 on People, Businesses, and Communities Using Private Sector Data’^[9], the researchers have built a tracker to measure economic activity at a granular level with the help of anonymized data from several private companies to construct indices of spending, employment, and other metrics and provide valuable information for understanding the state of the economy and facilitating the national recovery (<https://www.tracktherecovery.org/>).

The paper, The Global Macroeconomic Impacts of COVID-19: Seven Scenarios^[10] demonstrates the epidemiological scenarios in an economic model by creating a set of filters to convert the shocks into economic shocks to reduced labor supply in each country, rising cost of doing business in each sector including disruption of production networks in each country and also consumption reduction due to shifts in consumer preferences over each good from each country.

Now all these papers we have reviewed mostly covered the economic impacts in terms of business sector wise or w.r.t to GDP , where as ,our dataset is based on a survey responses and focused on individuals rather .Therefore the insights we have observed from our data-mining tasks focus on individual important attributes like age , monthly income , loan taken in order to maintain the expenses due to COVID-19, household responsibility and therefore are more people oriented instead of categorizing into multiple business sectors .

DATA

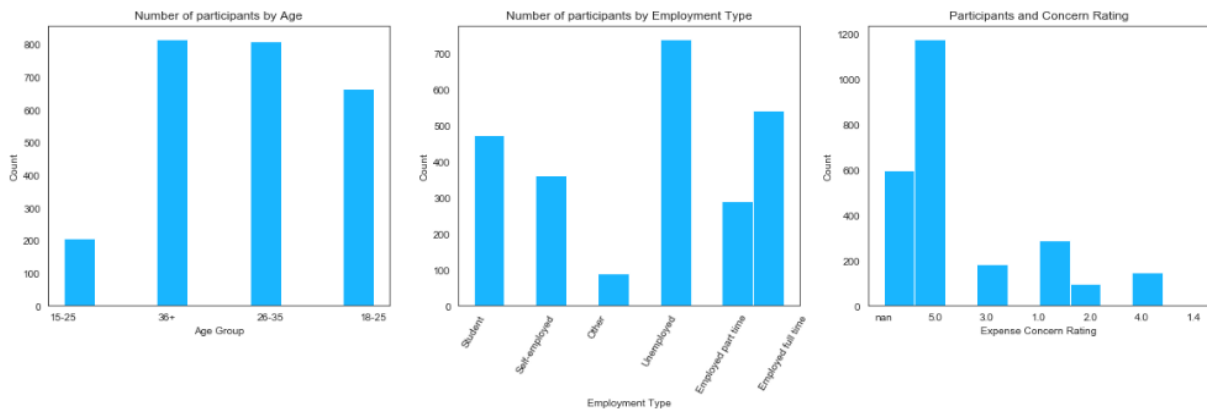
The data has been collected from the study implemented by GeoPoll using their own platform and respondent database. The questionnaire was designed by GeoPoll's questionnaire team. The questionnaire consisted of wide range of questions like demographic information, age, gender, employment type, type of work, expenses concern, priority of expenses, if aid received, source of aid, and the opinion on what should be the government priority. The study was conducted in different countries of the Sub-Saharan Africa. The study was conducted by SMS, where a two-way conversation takes place over the text messages. One survey question is sent at a time and the response to the survey is sent using short codes. The participants were selected randomly and hence the sample is representative of the country.^[11]

About the data

The dataset contains 2500 observations and 36 variables. It contains 500 observations each from the 5 participating countries. Equal numbers of males and females were surveyed.

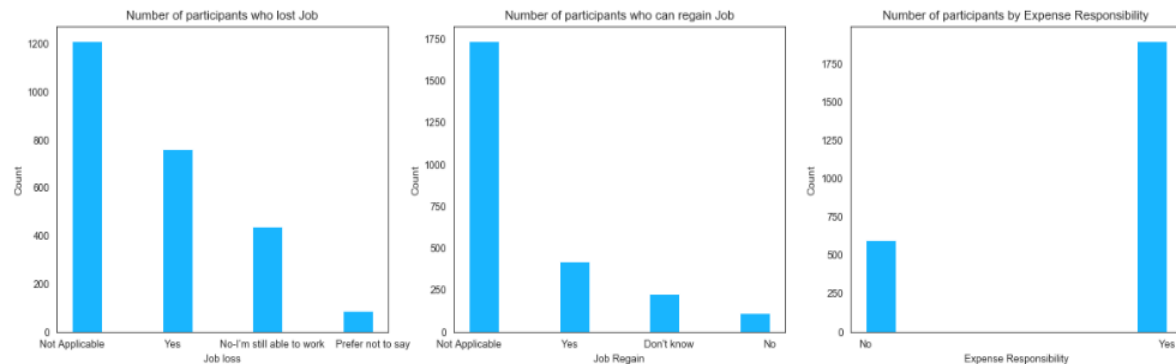
Age and BirthType are of data type 'int', MonthlyIncome Bracket, Express Concern rating are of data type 'float'. All the other variables are of data type 'object'. Detailed description of all the variable data types and number of non-null values can be found in appendix-data description.

Let us try to get some insights from the data which can help in choosing the variables for our model. As most our data is categorical, we plot some histograms to see how the data is distributed.

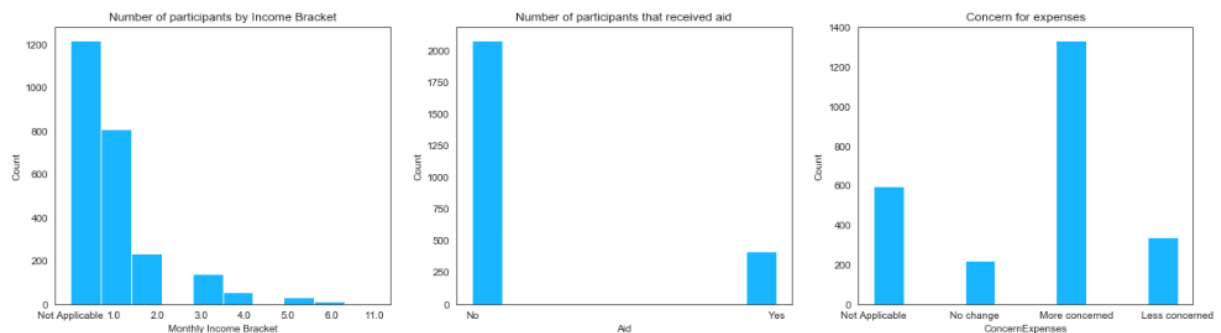


From the histogram of 'Age Group' above, we can observe that most of the participants are in the age group of 36+ and 26-35. This is generally the work group age and hence we may say that most of the participants belong to the work age group. From the histogram of 'Employment Type' we can observe that most of the participants are un-employed. This gives us the insight that there is high un-employment in African countries. The next category of participants in the survey are employed full-time. From the histogram of 'Express Concern Rating', we can observe that most of the participants have expressed their concern as '5' regarding expenses for the next

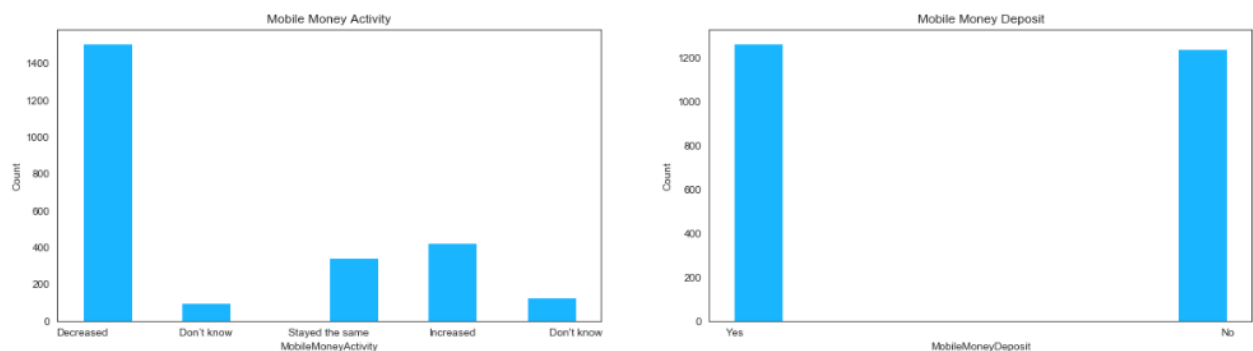
6 months. From this, we can say that people are highly concerned about the expenses due to COVID-19.



From the histogram of 'Job Loss' above, we can see that many people have lost jobs due to COVID-19. From the histogram of 'Job Regain', we can see that few people can regain job after COVID-19. From the histogram of 'Expense Responsibility' we can see that around 2000 people have stated that they have expense responsibility.



From the histogram of 'Monthly Income Bracket' we can say that most of the participants belong to the monthly income bracket of 1(which means they earn 0-10000). From the histogram of 'Aid' we can say that 2100 have not received any form of aid. From the histogram of 'Concern For Expenses' which gives the information about the concern of participant regarding expenses in for the next month, we can say that around 1400 people have stated that they are more concerned about expenses for the next month.



Mobile Money Activity is one of the important ways of carrying out transactions in the African countries. Hence, to know about the economic activity of the people of African countries it is important to study the Mobile money activity. From the histogram of mobile money activity, we can say that due to COVID-19, the mobile money activity of people has decreased drastically. Around 1450 people have stated that the mobile money activity has decreased due to COVID-19. If we look at the histogram of mobile money deposit during COVID-19, we can say that there are 50% of people who has deposited money in mobile money and 50% has not deposited any money into the mobile money during COVID-19.

Data cleaning and transformation

As the data has been collected from a survey with limited ways to answer, most of the variables had values properly entered and did not need to clean it further.

However, when we look at the column 'MonthlyNeed', we have observed that for few cases, different range of the values are joined by pipe. As this column will be one of the features in our next model building phase, we need to change the layout of these values before taking this as input feature. For that, first we have converted the column to string type and using the split() function of string, we will only pick the first bucket as for our future use. After performing this operation, when we checked the unique values of MonthlyNeed column, we can see that the buckets present in the column are not aligned with the survey questionnaire bucket value. To make that change, first we converted the 'Nan' values to the first bucket '0-10000'. This is considering that the Nan values are because people do not have a monthly need. Next, we have replaced the buckets – 'Over 1000000', 'Over 500001', '60001+' with their correct bucket as per the value in the survey questions. Finally we have defined a function **monthneedencode()** which will check each value of the column, retrieve the lower and upper bound value of each bucket and then by comparing them with the bucket values advised as per the survey questionnaire, will assign each of them to correct bucket. After applying this function to the MonthlyNeed column in our data frame, we will finally be able to get our correct bucket values.

For the 'LengthSurvival' and 'MobileMoneyActivity' columns, we have observed that the same value – 'Don't know' is been presented in two ways – it is possibly because of data-entry error, for some rows they are expressed with Don't know and for some they are populated as - Don't know – difference in terms of comma and single quote in the word 'Dont'. As they both represent the same values, we have replaced both the values to the name as 'Donot know' to maintain the uniformity.

We observed that there are few columns gives the duplicate information – as we have already column Age and Age Group in our dataset, we can drop the 'BirthYear' column from our analysis. Also as we are more interested whether the concerned person is receiving any AID or not and there were no missing values in that AID column, we can ignore the types of aid and therefore can drop the columns associated with different type of aids – as they represent if the

AID value is Yes , what is the corresponding type – (Government/religious organization/personal- family or friends). Similarly, as we already know that the dataset is from the South-Africa regions, we have dropped the Country, Province columns as we would like to focus on the results for the entire region rather than separating them by Country/Province. We are basically interested on the Employment type – which depict whether the person concerned is student/employed(full or part-time)/self-employed/unemployed so we have dropped the ‘JobType’, ‘InformalWorker’, ‘Informal Work Type’, ‘OtherJob’ columns from further considerations. As the ‘MonthlyIncome Bracket’ column already depicts the income range in terms of the corresponding bucket value, we also drop the ‘MonthlyIncome’ column as it represents the same information.

There are further transformations done to the variables to implement the model. This is explained clearly in the Methodology section.

Dealing with missing data

For the columns that have missing values, we have created a category and assigned that value to the missing values. This is done because removing all the observations with Nan values would result in loss of large amount of data.

For ‘MonthlyIncome Bracket’ column which is basically numeric value based on the ‘Monthly Income’ bucket range. There are total 1218 missing values. We have replaced the null values with the first bucket 1 as that represents the income range of (0-15000). This is because the ‘EmploymentType’ for the corresponding Nan values are either Student / Unemployed. Hence it makes sense to replace them with first bucket (considering this bucket included value = 0 as well).

For the columns ‘JobLoss’ which has 1212 missing values and ‘JobRegain’ which contain 1737 missing values each, we have replaced the null value as ‘Not Applicable’ This is done by considering the corresponding ‘EmploymentType’ for missing values in ‘JobLoss’ or ‘JobRegain’ is Student/Unemployed/Self-employed. Hence, JobLoss and JobRegain do not apply to these participants.

The same approach has been taken for ‘ConcernExpenses’ column. The column has 600 missing values. The missing values are for the observations who’s corresponding ‘ExpenseResponsibility’ has value No. Hence, we replace the missing values with ‘Not Applicable’ category.

The column ‘Expense Concern Rating’ has 600 missing values. The ‘ExpenseResponsibility’ column has value ‘No’ for corresponding missing values and hence we have replaced them with value 1(Not concerned). There is one value 1.4 and it is changed to 1 based on its income value. This could be a possible data entry error.

SUPERVISED LEARNING

The cleaning and transformation of dataset is completed, we now move on to the next phase, building a supervised learning algorithm to predict the target variable “Expense Concern Rating”

Here in the survey data, now we have the following information:- participants Age, Gender, Employment Type, Monthly Income, If Job lost due to COVID-19, Monthly Need, whether they are responsible for household expenses, currently they have loans or not, whether the person has a job to go back after the COVID restriction will be lifted, whether the income has been changed from initial months due to COVID-19, with the current income how long will the person expect to be able to pay for basic expenses like rent/food, as compared to before coronavirus, how concerned the person with paying the expenses in the current month, how will the person cover their expenses in the next month – by their salary /savings / loan/credit, whether the person has taken any AID , whether the person has taken any loans to cover their expenses due to COVID-19, how has the personal Mobile Money activity changed since COVID-19 and whether the person is able to deposit money or cash out at a Mobile Money agent since COVID-19 and what the person is thinking about the government’s priority at present. Now based on these variables, we want to estimate how much concerned will the person be with paying their expenses over the next 6 months of time on the scale of {1 : 'Not Concerned at all', 2 : 'Less concerned', 3:'Concerned', 4 : 'More concerned', 5 : 'Extremely concerned' }.

After careful consideration, as we want to predict the concern level of person for next six months and the levels are scaled from 1 to 5, we have decided to build classifier models^[12] as part of supervised learning task.

Data Preparation for classifier model

We have observed previously that most of the variables in our data are in categorical form, we need to address this first as most of the machine learning algorithms in Python (with scikit-learn) consider numerical values as input. First of all, we have addressed the columns which can be expressed as ordinal variables – for ‘ConcernExpenses’ column we have mapped the values to numeric values in a scale from 0 to 3 by using the map as ({'No change': 0,'Not concerned': 1 , 'Less concerned':2,'More concerned':3}). Similarly, we have created **monthlyneedmap** dictionary as {'0-10000': 1, '10001-20000':2, '20001-50000':3, '50001-100000':4, '100001-200000':5, 'Over 200000':6} and then applied it to ‘MonthlyNeed’ column to change the values to integer form.

For the target variable- ‘Expense Concern Rating’ column , we have created the **concernratingmap** in the form of {'Not Concerned at all': 1, 'Less concerned':2,'Concerned':3, 'More concerned':4, 'Extremely concerned':5} and then applied it to the column to change the values to integer format.

For the rest of the variables 'Gender', 'EmploymentType', 'JobLoss', 'JobRegain', 'IncomeChange', 'ExpenseResponsibility', 'LengthSurvival', 'MoneyForExpenses', 'Aid', 'COVIDLoans', 'MobileMoneyActivity', 'MobileMoneyDeposit', 'GovernmentPriority' which have data type as 'object', we will convert each category value of these columns into a new column and assign a 1 or 0 (True/False) value to the column by using `get_dummies()`. Finally, we will divide the data frame into independent variables (the one from which we will predict the target) and the dependent or predicted variable i.e. 'Expense Concern Rating' for classification task.

Initial Decision-Tree Classifier Model

For our initial tree-based model, first we have created a Pipeline – with preprocessor and classifier as it's steps. As part of preprocessor, from the original dataset, we filter the columns with object type and will encode them by one hot encoding and then in the classifier step, a Decision Tree classifier will be created on this processed dataset. The advantage of using this Pipeline is that we do not have to perform any transformation for any column of the data frame when passing into the model, rather it will handle automatically the categorical variables and leave the numeric variables unchanged before passing it to the model. [**Note:** We have performed this model before applying the steps as mentioned in the previous section of Data Preparation for Model Building. The idea here was to get the idea of our first model]. With our initial model, we have received the **accuracy score of 68.5 %**. One thing which we have observed that instead of handling all the object type variable with One-Hot encoding, it will be more useful to address some of the columns which can be ordinal in nature separately and that was our motivation to all the steps in the previous step where we create 'monthlyneedmap' and 'concernratingmap' to address some columns like 'ConcernExpenses', 'MonthlyNeed' and 'Expense Concern Rating' separately.

After performing the data preprocessing steps (few columns treated individually) and dividing the dataset into independent variables and dependent variable, we split the datasets into 67% training -> 1675 samples and 33% test data -> 825 samples. Then we train a Decision Tree classifier with criterion = as 'entropy', max_depth =5 and min_samples_split=100 and then we predict the Expense Concern Rating value from the test dataset .Finally to evaluate the classifier , we have verified the accuracy score and this model gives us **70% accuracy score on test data**.

First Decision Tree

```

graph TD
    Root["ExpenseResponsibility_Yes <= 0.5  
entropy = 1.75  
samples = 1975  
value = [87, 63, 125, 108, 794]  
class = 5"]
    Root -- True --> Leaf1["entropy = 0.0  
samples = 398  
value = [568, 0, 0, 0]  
class = 1"]
    Root -- False --> Node1["ConcernExpenses <= 2.5  
entropy = 1.675  
samples = 1277  
value = [189, 83, 125, 108, 794]  
class = 5"]
    Node1 --> Node2["MonthlyNeed <= 5.5  
entropy = 1.463  
samples = 383  
value = [81, 35, 57, 26, 185]  
class = 5"]
    Node1 --> Node3["Aid_No <= 0.5  
entropy = 1.469  
samples = 894  
value = [108, 23, 68, 81, 601]  
class = 5"]
    Node2 --> Node4["JobRegain_Don't know <= 0.5  
entropy = 2.025  
samples = 153  
value = [30, 12, 37, 13, 71]  
class = 5"]
    Node2 --> Node5["MobileMoneyActivity_Deceased <= 0.5  
entropy = 1.864  
samples = 220  
value = [81, 23, 20, 12, 114]  
class = 5"]
    Node3 --> Node6["LengthSurvival_Donot know <= 0.5  
entropy = 1.883  
samples = 145  
value = [26, 6, 13, 23, 77]  
class = 5"]
    Node3 --> Node7["LengthSurvival_1 month <= 0.5  
entropy = 1.412  
samples = 743  
value = [82, 22, 56, 58, 53]  
class = 5"]
    Node4 --> Leaf2["LengthSurvival_1 month <= 0.5  
entropy = 2.024  
samples = 150  
value = [27, 12, 30, 13, 08]  
class = 5"]
    Node4 --> Leaf3["entropy = 1.457  
samples = 13  
value = [5, 0, 7, 0, 3]  
class = 3"]
    Node5 --> Leaf4["entropy = 1.863  
samples = 99  
value = [21, 5, 12, 9, 52]  
class = 5"]
    Node5 --> Leaf5["MoneyForExpenses_Other <= 0.5  
entropy = 1.793  
samples = 121  
value = [30, 18, 8, 3, 62]  
class = 5"]
    Node6 --> Leaf6["JobRegain_Not Applicable <= 0.5  
entropy = 1.895  
samples = 119  
value = [25, 5, 13, 15, 61]  
class = 5"]
    Node6 --> Leaf7["entropy = 1.316  
samples = 26  
value = [1, 1, 0, 8, 16]  
class = 5"]
    Node7 --> Leaf8["MonthlyNeed <= 4.0  
entropy = 1.311  
samples = 623  
value = [58, 16, 43, 44, 482]  
class = 5"]
    Node7 --> Leaf9["Age <= 39.5  
entropy = 1.811  
samples = 128  
value = [24, 6, 12, 14, 70]  
class = 5"]
    Leaf2 --> Leaf10["entropy = 1.968  
samples = 123  
value = [18, 12, 23, 10, 60]  
class = 5"]
    Leaf2 --> Leaf11["entropy = 1.906  
samples = 27  
value = [0, 0, 7, 3, 8]  
class = 5"]
    Leaf5 --> Leaf12["entropy = 1.766  
samples = 15  
value = [30, 17, 5, 2, 51]  
class = 5"]
    Leaf5 --> Leaf13["entropy = 1.324  
samples = 16  
value = [0, 1, 3, 1, 11]  
class = 5"]
    Leaf7 --> Leaf14["entropy = 1.591  
samples = 40  
value = [4, 0, 3, 3, 20]  
class = 5"]
    Leaf7 --> Leaf15["entropy = 1.93  
samples = 79  
value = [1, 1, 5, 10, 12, 41]  
class = 5"]
    Leaf8 --> Leaf16["entropy = 1.322  
samples = 188  
value = [0, 4, 21, 19, 143]  
class = 5"]
    Leaf8 --> Leaf17["entropy = 1.378  
samples = 427  
value = [40, 12, 22, 25, 318]  
class = 5"]
    Leaf9 --> Leaf18["entropy = 1.8  
samples = 102  
value = [14, 6, 12, 11, 59]  
class = 5"]
    Leaf9 --> Leaf19["entropy = 1.417  
samples = 24  
value = [15, 0, 0, 3, 11]  
class = 5"]
  
```

accuracy to 71.2%. We can then experiment this tree by tuning its hyperparameters with two methods:

(a) **RandomizedSearchCV Method**

We use this method to train and evaluate a series of models by taking random draws from a predetermined set of hyperparameter distributions and outputs a model which is trained on optimal set of hyperparameters. Here we have selected these parameters - 'n_estimators', 'max_features', 'max_depth', 'min_samples_split', 'min_samples_leaf', 'bootstrap' for tuning purpose and provided the range of values from which the algorithm can select on. After fitting 5 folds for each of 100 candidates, total of 500 fits, we have received the best parameters for this Random Forest Classifier model and with the best selected parameters, the **accuracy score has improved to 73.6%.**

(b) **GridSearchCV Method**

Like our earlier decision tree classifier, here also we will create a grid of parameters and their range of values for Random Forest Classifier as follows:

```
gridsearch_parameterer = {  
    'n_estimators': [150, 600],  
    'max_features': ['auto', 'sqrt', 'log2'],  
    'max_depth': [4,5,6,7,8, None],  
    'criterion': ['gini', 'entropy'],  
    'bootstrap': [True]  
}
```

Now with these best estimated parameters, we have observed the **accuracy score as 71.4%.** Hence, **RandomizedSearchCV** process is providing the **best accuracy score (73.6%)** on Random Forest Classifier model so far.

Other Classifier Models

Apart from the tree-based classifier models, we have also experimented with other classifier models – **Gaussian Naïve Bayes** and **Logistic Classifier for multiclass**. Now, for the **Gaussian Naïve Bayes classifier**, we have received a very **poor accuracy score(42%)**. One of the possible reasoning may be as in our dataset, we have many categorical variables which have been one-hot encoded-therefore they are mainly containing either zeros or ones. In case of Gaussian Naïve Bayes, it models all these values following a normal distribution- but as we know most of the variables are only zeroes and ones, we may have losing important information in our model.

Next, we have trained our data on **Logistic Multiclass Classifier** by using one Vs rest approach (with multi_class='ovr'). Before training our model, we have also standardized our data with the help of StandardScaler() method and the **accuracy is 71%** and **238 records** have been classified

incorrectly. So far, Random Forest Classifier still has higher accuracy score so we will model our data based on that.

Feature Selection

As our dataset has many independent variables, feature selection can play important role as we can make our model simpler

to interpret, also selecting only important relevant features will reduce the variance of our model and therefore will also be less overfitting. Now for the Random Forest model, it naturally ranks by how well they improve the purity of the node – the nodes with the greatest decrease in impurity generally occur at the beginning of the trees and the ones with least decrease in impurity happen at the bottom of the tree. Therefore, we can select features by using the feature ranks generated by these as a ranking, and then can prune the features based on that ranking. Here we have used **SelectModel()**, which will use the Random Forest Classifier to identify the important features (The features which have an importance more than mean of the overall importance score of all feature). Then we can use those important features (retrieved by **get_support()** method) and create a subset of our dataset containing only these important features. Finally, we will train a new Random Forest Classifier with only these important features and predict our Expense Concern rating column. We can see that **SelectModel()** method lists the **features as most important** : 'Age', 'ConcernExpenses', 'MonthlyNeed', 'ExpenseResponsibility_No', 'ExpenseResponsibility_Yes', 'LengthSurvival_Not Applicable', 'MoneyForExpenses_Not Applicable'. With only these important features and the best estimated hyperparameters as obtained previously, when we train a new Random Forest classifier, we have obtained the **accuracy score of 68.7% on test-data**. Now, as we can see that we have come up with the **reduced feature dimensionality** at the expense of the **reduced accuracy score from 73% to 68.7%**.

Secondly, we have used the **recursive feature elimination (RFE)** method which is an optimization algorithm. This algorithm repeatedly creates models and keeps aside the best or the worst performing feature at each iteration. It constructs the next model with the remaining a feature and continue until all the features are exhausted. It then ranks the features based on the order of their elimination. Using the features listed by this RFE method and with the best estimated hyperparameters, we have trained a new classifier and achieved the **accuracy score as 70%** with the **12 features** as follows: 'Age', 'MonthlyIncome Bracket', 'ConcernExpenses', 'MonthlyNeed', 'ExpenseResponsibility_No', 'Aid_Yes', 'ExpenseResponsibility_Yes', 'LengthSurvival_< a month', 'LengthSurvival_Not Applicable', 'MoneyForExpenses_Loan/ Credit', 'MoneyForExpenses_Not Applicable', , 'MobileMoneyDeposit_Yes'

Summary of all the models and features used can be found in Appendix-Model Summary

UNSUPERVISED LEARNING

We perform the Apriori association rule mining algorithm to generate frequent item sets with a threshold support and confidence on this dataset. The algorithm generates association rules between various kinds of social and economic factors.

After dropping the columns that contain similar information, we are left with the columns:

Age Group, Gender, EmploymentType, JobLoss, JobRegain, IncomeChange, Expense Responsibility, LengthSurvival, MoneyForExpenses, ConcernExpenses, Aid, COVIDLoans, MobileMoneyActivity, MobilemoneyDeposit, GovernmentPriority. These are all of data type 'object'. 'Expense Concern Rating', 'MonthlyIncome Bracket' of type float.

Including all variables

We begin by including all the above variables and generate the association rules.

Data Preparation for Apriori algorithm

The N/A values in JobLoss are replaced by 'Job Loss-Not Applicable'. We have transformed the other values as: No-I'm still able to work: 'Job Not lost due to Covid', Yes : 'Job Lost due to Covid', Prefer not to say : 'Prefer not stay about Job loss'.

The N/A values in JobRegain are replaced by 'Job Regain-Not Applicable', The other values are transformed as : Don't know : 'Not sure about job regain', Yes : 'Can regain job', No : 'Cannot regain job'.

The N/A value of variable 'MonthlyIncome Bracket' are replaced with income bracket 1. There is one value '11' but the corresponding monthly income is 0-5000 and hence it is added to bracket 1 considering it a data-error. The variable is converted to type string and other values are transformed as : 1 : 'Income between 0-10000', 2 : 'Income between 10001-20000', 3 : 'Income between 20001-50000', 4 : 'Income between 50001-100000', 5 : 'Income between 100001-200000', 6- Income above 200000.

The variable 'IncomeChange' has 1213 missing values. All these missing values are replaced with the value 'Income Change-Not Applicable'. There are many unemployed and student participants in the survey and 'IncomeChange' question does not apply to them. Hence, we have chosen to replace N/A. In addition, the other values are transformed as below:

Decreased a bit : 'Income change decreased', Decreased a lot : 'Income change decreased', Increased a bit : 'Income change increased', increased a lot : 'Income change increased', No change : 'No change in income' . Here, the number of levels of the variable are reduced to 4 levels from 7.

The variable 'ExpenseResponsibility' values are transformed as : No : 'No Expense Responsibility', Yes : 'Has Expense Responsibility'.

The variable 'LengthSurvival' has 600 missing values. They have been replaced with 'Unsure about length survival'. The other values are replaced as below:

Don't know : Unsure about length survival, 5+ months : Can survive for 5+ months, 2-4 months : Can survive for 0-4 months, 1 month : Can Survive for 0-4 months, < a month : Can survive for 0-4 months. Here, the number of levels of the variable are reduced to 3 levels from 5.

The variable 'MoneyForExpenses' has 600 missing values. The corresponding 'EmploymentType' values for the missing variables are student/unemployed/self-employed and 'ExpenseResponsibility' is 'No'. Hence the missing values are replaced with 'Other way for expenses'.

The variable 'ConcernExpenses' has 600 missing values. The corresponding 'EmploymentType' values for the missing variables are student/unemployed/self-employed and 'ExpenseResponsibility' is 'No'. Hence the missing values are replaced with 'No change in concern for expenses'. The other values are transformed as : Less concerned : Less concerned about expenses than before covid, More concerned : More concerned about expenses than before covid, No change : No change in concern for expenses.

The variable 'Expense Concern Rating' has 600 missing values. The variable is converted to type string. The corresponding 'EmploymentType' values for the missing variables are student/unemployed/self-employed and 'ExpenseResponsibility' is 'No'. Hence the missing values are replaced with 'Not concerned for next 6 months'. The other values are transformed as: 1 : Not concerned for next 6 months, 2.0 : Reasonably concerned for next 6 months, 3.0 : Reasonably concerned for next 6 months, 4.0 : Reasonably concerned for next 6 months, 5.0 : Extremely concerned for next 6 months. The number of levels reduced from 6 to 3.

The 'Aid' column values are transformed as: No : Aid Not Received, Yes : Aid Received.

The 'COVIDLoan' column values are transformed as: No : No Loans taken due to Covid, Yes : Loans taken due to Covid.

The 'MobileMoneyActivity' column values are transformed as : Decreased : Mobile Money activity Decreased, Increased : Mobile Money activity Increased, Stayed the same : Mobile Money activity stayed the same, Don't know : Unsure of Mobile Money Activity.

The 'MobileMoneyDeposit' column values are transformed as : No : No mobile money deposit after Covid, Yes : There is mobile money deposit after Covid.

The 'GovernmentPriority' column is dropped because this column just gives the opinion of the participant and not any current situation. Hence, this need not be considered for the algorithm.

Apriori Algorithm

To implement the Apriori^[14] algorithm, we need to convert the data into the transactional format. We first convert the data frame into numpy array. Now using **TransactionEncoder()** function, we format the data into transactional format. The values are transformed into columns and the corresponding value in the column for the observation is True when the observation is of that level or else False. Now we have the data transformed into 57 columns and 2500 observations. Now as we have all the data ready, we use **apriori()** to generate frequent item sets. It takes the parameters of data, min_support. Min_support is the threshold and all the item sets above the threshold are displayed.

Once the frequent item sets are generated, we use **association_rules()** to generate the association rules. To this function we pass the frequent item sets output as data as we need to generate rules that support the threshold level given. The metric is chosen as 'confidence' and min_threshold parameter is specified. All the rules above this threshold are generated.

Frequent item sets and rules generated with different support and confidence levels are:

1. Support -50%, Confidence – 50%

We have 17 frequent item sets and 20 rules generated. The support of the itemset range from 50% to 83%. The confidence of the rules ranges from 50% to 100%. All though there are many rules generated the rules are strongly associated only when lift is > 1 ^[15]. There are 2 rules for which lift is > 1.1 . They are:

{More concerned about expenses than before COVID-19 <-> Has expense responsibility}

The above rule occurs with 100% confidence. So, we can interpret that if people have expense responsibility, they are more concerned about expenses. There are 10 other rules with lift value slightly greater than 1.

2. Support - 60%, Confidence - 60%

There are 8 frequent item sets and 6 rules generated. There are no rules with lift > 1.1 . hence, we can say that these rules are not strongly associated with each other. There are 4 rules with lift value slightly greater than 1.

3. Support – 70%, Confidence – 70%

There are 3 frequent itemset generated but no rules with minimum confidence 70%.

Interpreting above rules

Although we have generated many rules by considering all the variables, most of the rules are generated are from the Not Applicable category. This happens because for the variables that contain missing values, most observations fall under the Not Applicable category.

Though these rules have high confidence values there is no meaningful inference from these rules. There are no meaningful patterns that can be deduced from the rules that contain Not Applicable either as an antecedent or consequent. From the one rule that has lift value of 1.3 we can deduce that if a person has expense responsibility then the person is more concerned about expenses than before COVID-19.

Removing Not Applicable values and generating rules

Using N/A values in our dataset is creating ambiguity in the rules and hence we remove the N/A values and try to find association between the columns. We chose all the same columns as above and drop other columns. Now remove all the rows that has N/A values. If any variable has value N/A then the corresponding row is deleted. After this transformation there are 661 observations and 16 columns.

We use the apriori algorithm to generate the rules. We transform the data into transactional format and perform the algorithm with different support and threshold levels.

Frequent itemset and rules generated with different support and confidence levels are:

1. Support - 60%, Confidence - 60%
There are 34 frequent item sets and 118 rules generated. The lift value for most of the rules is 1.
Let us reduce the count of the rules by increasing the support and confidence values.
2. Support – 70%, Confidence – 70%
There are 15 frequent itemset and 32 rules generated with lift value 1.0 for all the rules.
This implies that there is no strong association between the rules.
3. Support – 80%, Confidence – 80%
There are 7 frequent item sets and 12 rules generated with lift value 1.0 implying that there is no strong association between the rules.
4. Support – 90%, Confidence – 90%
There are 3 frequent itemset and 2 rules generated with lift value 1.0.

Interpreting above rules

The rules that are generated are generated with confidence and support >60%. Although we can see high support and confidence, the lift value for all the rules is 1.0 implying that the antecedents and the consequents are independent of each other and do not exhibit strong association.

Hence, the accuracy of the rules is very low.

But we can draw some patterns from the frequent item sets and the confidence of the rules. 'Aid Not Received' has high support value and the rule Aid not received <-> Has expense responsibility has high confidence value.

Another rule Has Expense Responsibility <-> job lost due to Covid-19 has high confidence

Another rule Aid not received <-> job lost due to Covid-19 has high confidence.

From this we can interpret that people who has expense responsibility, has lost job due to COVID-19 , has not received any aid are more frequent to occur. Hence, the government could use these patterns to identify the people who has lost job due to COVID-19 with expense responsibility and provide some aid to them.

Using important features and generating rules

We use important features that are generated from supervised learning to generate rules. We consider the columns 'MonthlyIncome Bracket', 'Concern Expenses', 'MonthlyNeed', 'ExpenseResponsibility', 'LengthSurvival', 'MoneyForExpenses', 'Express Concern Rating'. We drop the observations that have missing values in any of the columns. Now, We are left with 1091 observations and 6 columns We use the apriori algorithm to generate the rules. We transform the data into transactional format and perform the algorithm with different support and threshold levels.

Frequent itemset and rules generated with different support and confidence levels are:

1. Support = 50%, Confidence = 50%
There are 9 frequent item sets and 8 rules generated. The lift value for all the rules is 1.0.
2. Support = 60%, Confidence = 60%
There are 5 frequent itemset and 4 rules generated. The lift value for all the rules is 1.0.
3. Support = 70%, Confidence = 70%
There is 1 frequent itemset and 0 rules generated.

Interpreting the results

Lift value for all the rules is 1.0 implying that the antecedents and the consequents are independent of each other and do not exhibit strong association.

Hence, the accuracy of the rules is very low. Due to the low accuracy of the rules we cannot say the occurrence of antecedent will also lead to the occurrence of consequent. But we can see some patterns in the rules that are generated with high support and confidence.

People who has expense responsibility <-> can survive for 0-4 months. People who has expense responsibility <-> Extremely concerned for next 6 months. People whose income between 0-1000 <-> Has expense responsibility. These rules occur with 100% confidence. Government can combine these 3 rules and find out the people who has income between 0-1000 and with expense responsibility and provide them with some aid to reduce the express concern about expenses for next 6 months.

CONCLUSION

This report analyzes the economic impact of COVID-19 on people of Sub-Saharan Africa based on a survey dataset published on “HUMANITARIAN DATA EXCHANGE” site. The models from our dataset highlights the issue that most of the people in Africa are highly concerned about their expenses for upcoming six months due to pandemic. The important features which are attributed in the estimation of the expense concern are age, monthly income, monthly need, responsibility for keeping the household expense. Most of the people who expressed their high concern are maintaining their expenses based on loan or credit.

From the association rule mining, we have observed that people responsible for their household expenses, have lower monthly income (fall in the first bracket who earn between 0-10000), are expected to be able to pay the basic expense of food and rent for fewer months from now with their current income, have received no aid from any organization and are highly concerned about their expenses for next 6 months.

In terms of the data-mining process, we have observed that while building classifier model on our dataset, tree-based model works best. Random Forest classifier provides better model accuracy score than regular Decision-Tree. While using the feature selection method, we have received lower accuracy score compared to original Random Forest with almost all the features but at the same time we have reduced the dimensionality (12 features) in our model which is more easy to interpret and prone to less overfitting. Logistic Multiclass model performs moderately well on our dataset, but Gaussian Naïve Bayes has very poor performance and one of the possible reasons can be the large number of categorical values which we have encoded by dummy variables.

While performing apriori algorithm to generate rules we have found that the rules generated has changed with the change in dimensionality of the data. Although the rules that are generated have reasonable support and confidence values (around 80%), the lift value of the rules is 1.0 which basically says that our rules are not strongly associated. Hence, rather than using the findings as rules we can use them as patterns observed.

Finally, the prediction results and the important features attributed to the estimation will be helpful for the Government of Africa to determine how COVID-19 has impacted the common people financially – how much they are concerned in continuing their expenses during this pandemic. The insights from the different association rule mining will attribute to construct the economic rebuilding roadmaps in the post pandemic time and provide the informative features to be used by the Government of Africa for their different aid programs targeting the correct people.

APPENDIX

References

- [1] The Financial Impact of Coronavirus in Sub-Saharan Africa. GeoPoll Report.
https://f.hubspotusercontent30.net/hubfs/325431/Financial%20Impact%20of%20COVID%20Report_Final.pdf?utm_campaign=Financial%20Impact%20Report&utm_source=GeoPollBlog&utm_medium=imagelink
- [2] Ataguba, J. E. (2020). COVID-19 pandemic, a war to be won: understanding its economic implications for Africa.
- [3] Economic Commission for Africa (2020). COVID-19 in Africa: Protecting Lives and Economies.
- [4] Baker, S. R., Bloom, N., Davis, S. J., & Terry, S. J. (2020). *Covid-induced economic uncertainty* (No. w26983). National Bureau of Economic Research.
- [5] Evelina, J., Samuel, N., & Lukas Homateni, J. (2020). Estimating the Economic Impact of COVID-19: A Case Study of Namibia.
- [6] Barua, S. (2020). Understanding Coronanomics: The economic implications of the coronavirus (COVID-19) pandemic. *SSRN Electronic Journal* <https://doi.org/10/ggq92n>.
- [7] Kassem, J. J. (2020). COVID-19 Outbreak: Is It a Health Crisis or Economic Crisis or Both? Case of African Counties. *Case of African Counties (March 23, 2020)*.
- [8] Arndt, C., Davies, R., Gabriel, S., Harris, L., Makrelov, K., Modise, B., ... & Anderson, L. (2020). Impact of Covid-19 on the South African economy. *Southern Africa-Towards Inclusive Economic Development Working Paper, 111*.
- [9] Chetty, R., Friedman, J. N., Hendren, N., & Stepner, M. (2020). Real-time economics: A new platform to track the impacts of COVID-19 on people, businesses, and communities using private sector data. *NBER Working Paper, 27431*.
- [10] McKibbin, W. J., & Fernando, R. (2020). The global macroeconomic impacts of COVID-19: Seven scenarios.
- [11] Economic impact of COVID-19 in Sub-Saharan Africa (2020). GeoPoll Survey data dataset – Public data.
<https://data.humdata.org/dataset/economic-impact-of-covid-19-in-sub-saharan-africa>
- [12] Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., ... & Vanderplas, J. (2011). Scikit-learn: Machine learning in Python. *the Journal of machine Learning research, 12*, 2825-2830.
- [13] Peter Worcester.,(Jun 5, 2019). A comparison of Grid Search and Randomized Search using Scikit Learn. *Noteworthy-The Journal Blog*
<https://blog.usejournal.com/a-comparison-of-grid-search-and-randomized-search-using-scikit-learn-29823179bc85>
- [14] Sebastian Raschka. (2014-2020). Frequent Itemsets Via Apriori Algorithm
http://rasbt.github.io/mlxtend/user_guide/frequent_patterns/apriori/
- [15] Anisha Garg.(Sep 3, 2018). Complete guide to Association Rules(1/2). *towards data science*
<https://towardsdatascience.com/association-rules-2-aa9a77241654>

Data overview

0	Country	2500 non-null	object
1	Province	2500 non-null	object
2	Age Group	2500 non-null	object
3	Age	2500 non-null	int64
4	BirthYear	2500 non-null	int64
5	Gender	2500 non-null	object
6	EmploymentType	2500 non-null	object
7	JobType	836 non-null	object
8	InformalWorker	1227 non-null	object
9	Informal Work Type	642 non-null	object
10	OtherJob	417 non-null	object
11	JobLoss	1288 non-null	object
12	JobRegain	763 non-null	object
13	MonthlyIncome	1288 non-null	object
14	MonthlyIncome Bracket	1282 non-null	float64
15	IncomeChange	1287 non-null	object
16	ExpenseResponsibility	2500 non-null	object
17	LengthSurvival	1900 non-null	object
18	MoneyForExpenses	1900 non-null	object
19	ConcernExpenses	1900 non-null	object
20	Expense Concern Rating	1900 non-null	float64
21	MonthlyNeed	1932 non-null	object
22	TopPriority	1968 non-null	object
23	LowPriority	1968 non-null	object
24	Aid	2500 non-null	object
25	AidSource_Aid Organisations	419 non-null	object
26	AidSource_Charities/ Donations	419 non-null	object
27	AidSource_Friends/family	419 non-null	object
28	AidSource_Government	419 non-null	object
29	AidSource_Not sure	419 non-null	object
30	AidSource_Other	419 non-null	object
31	AidSource_Religious Organizations	419 non-null	object
32	COVIDLoans	2500 non-null	object
33	MobileMoneyActivity	2500 non-null	object
34	MobileMoneyDeposit	2500 non-null	object
35	GovernmentPriority	2500 non-null	object

Summary of the models used in supervised learning

Model	Features Used	Accuracy Score
Decision tree Classifier (Using Pipeline for handling the preprocessing with One -Hot encoding)	Gender', 'EmploymentType', 'JobLoss', 'JobRegain', 'IncomeChange', 'ExpenseResponsibility', 'LengthSurvival', 'MoneyForExpenses', 'Aid', 'COVIDLoans', 'MobileMoneyActivity', 'MobileMoneyDeposit', 'GovernmentPriority', 'ConcernExpenses', 'MonthlyNeed'	68.50%
Decision tree Classifier (Handle the preprocessing columnwise - ordinal data differently and rest with dummy variables)	Gender', 'EmploymentType', 'JobLoss', 'JobRegain', 'IncomeChange', 'ExpenseResponsibility', 'LengthSurvival', 'MoneyForExpenses', 'Aid', 'COVIDLoans', 'MobileMoneyActivity', 'MobileMoneyDeposit', 'GovernmentPriority', 'ConcernExpenses', 'MonthlyNeed'	70.70%
Decision Tree with selected features by SelectModel and using	Age', 'MonthlyIncome Bracket', 'ConcernExpenses', 'MonthlyNeed', 'ExpenseResponsibility_No', 'GovernmentPriority_Protecting people from COVID-19'	58.70%
Random Forest Tree Classifier	Gender', 'EmploymentType', 'JobLoss', 'JobRegain', 'IncomeChange', 'ExpenseResponsibility', 'LengthSurvival', 'MoneyForExpenses', 'Aid', 'COVIDLoans', 'MobileMoneyActivity', 'MobileMoneyDeposit', 'GovernmentPriority', 'ConcernExpenses', 'MonthlyNeed'	71.20%
Random Forest Tree Classifier with hyperparameters estimation by RandomizedSearchCV	Gender', 'EmploymentType', 'JobLoss', 'JobRegain', 'IncomeChange', 'ExpenseResponsibility', 'LengthSurvival', 'MoneyForExpenses', 'Aid', 'COVIDLoans', 'MobileMoneyActivity', 'MobileMoneyDeposit', 'GovernmentPriority', 'ConcernExpenses', 'MonthlyNeed'	73.60%
Random Forest Tree Classifier with hyperparameters estimation by GridSearch	Gender', 'EmploymentType', 'JobLoss', 'JobRegain', 'IncomeChange', 'ExpenseResponsibility', 'LengthSurvival', 'MoneyForExpenses', 'Aid', 'COVIDLoans', 'MobileMoneyActivity', 'MobileMoneyDeposit', 'GovernmentPriority', 'ConcernExpenses', 'MonthlyNeed'	71.20%
Random Forest Tree Classifier with features selected by SelectModel	Age', 'ConcernExpenses', 'MonthlyNeed', 'ExpenseResponsibility_No', 'ExpenseResponsibility_Yes', 'LengthSurvival_Not Applicable', 'MoneyForExpenses_Not Applicable'	68.70%
Random Forest Tree Classifier with features selected by Recursive Feature Elimination	Age', 'MonthlyIncome Bracket', 'ConcernExpenses', 'MonthlyNeed', 'ExpenseResponsibility_No', 'ExpenseResponsibility_Yes', 'LengthSurvival_< a month', 'LengthSurvival_Not Applicable', 'MoneyForExpenses_Loan/ Credit', 'MoneyForExpenses_Not Applicable', 'Aid_Yes', 'MobileMoneyDeposit_Yes'	70.20%
Gaussian NB Classifier	Gender', 'EmploymentType', 'JobLoss', 'JobRegain', 'IncomeChange', 'ExpenseResponsibility', 'LengthSurvival', 'MoneyForExpenses', 'Aid', 'COVIDLoans', 'MobileMoneyActivity', 'MobileMoneyDeposit', 'GovernmentPriority', 'ConcernExpenses', 'MonthlyNeed'	42.20%
Logistic Multiclass Classifier(One vs Rest)	Gender', 'EmploymentType', 'JobLoss', 'JobRegain', 'IncomeChange', 'ExpenseResponsibility', 'LengthSurvival', 'MoneyForExpenses', 'Aid', 'COVIDLoans', 'MobileMoneyActivity', 'MobileMoneyDeposit', 'GovernmentPriority', 'ConcernExpenses', 'MonthlyNeed'	71%