Global Scientific JOURNALS

DATA MINING AND ANALYSIS: UNDERSTANDING COVID-19 THROUGH ANALYSIS OF HEALTHY DIET DATASET

Bob Ntakirutimana[1] and Dr. Papias Niyigena[1]

[1]Faculty of Computing and Information Sciences, University of Lay Adventists of Kigali

Abstract

In the early stage of the COVID-19 outbreak, due to the limited data accumulated, the role of big data analysis was limited, hence the data mining models could help policymakers and health managers in private and public bodies to plan health care resources and control the prevention of an epidemic outbreak. This could be achieved by extracting data from different spots such as the media, social media networks used by service users and their relatives. The data acquired by Data Mining can likewise be used for publicizing, extortion recognition and logical discoveries and so on. To carry out this work and achieve its goal, which is to respond on the given question "What do we know about non-pharmaceutical interventions in fighting with coronavirus pandemic?" through data visualization.

Keywords: COVID-19, Data visualization, Healthy diet, Data mining

INTRODUCTION

With the disease spread worldwide, the World Health Organization (WHO) designated it as a pandemic by the beginning of 2020. Beyond the personal calamities and losses brought by this pandemic, the economic implications of this pandemic are significant (Abdelrahman, E. E. E. et al, 2020). The pandemic has brought a shadow over the entire world and an uncontrolled changes in what is trendy concerning the environment, society, food, and even individuals' general health (Shams, M. Y. et al, 2021).

The severe critical respiratory syndrome (SARS) that occurred in 2003 and the new type of corona pneumonia virus (COVID-19 or SARS-CoV-2) is an example of major public health

incidents, the spread of the new pandemic has led to a longer period of cessation of global economic activity, a greater demand shock, resulting in severe damage to human health and societal stability such as diet and health services (Qiong, J. et al, 2020); Though many studies on COVID-19 have recently emerged, and various data science applications combating the pandemic have been reported recently involved to find the data driven decisions to fight not only against the spread of the pandemic but even on the livings of population via the providing of healthy diet.

Violent disasters such as the COVID-19 pandemic can disrupt the usual food system and change relationship between human with food. Furthermore, in an effort to reduce the spread of infection, serious measures were taken where border and other logistic restrictions limiting the on-going of goods and people which increased the inadequate availability of necessity on market and lead to the risk of food shortages due to compromised supply chains, including those related to labor scarcities.

Good nutrition is very important before during and after an infection; Infections or disease such as Coronavirus take a toll on the body especially when these cause fever, the body needs extra energy and nutrients. While no foods or dietary supplements can prevent COVID-19 infection, maintaining a healthy diet is an important part of supporting a strong immune system; Diets vary greatly from place to place based on many factors including eating habits and culture. Yet, when it comes to food, there is a lot that we know about how to select the right combination of food to attain a healthy diet regardless of where we live (FAO, 2020).

In addition, various dietary ingredients are determinants of gut microbial composition and subsequently shape the immune responses in the body; certain factors such as lifestyle, age, health status, sex, and medications affect the nutritional status of an individual. Therefore, the existing evidence suggests that the only sustainable way to survive in the current situation is to strengthen the immune system (Faseeha, A. and Sadia, M., 2020).

In the early stage of the outbreak, due to the limited data accumulated, the role of big data analysis was limited, hence the data mining models could help policymakers and health managers in private and public bodies to plan health care resources and control the prevention of an epidemic outbreak. The availability of high-quality and timely data in the early stages of the outbreak collaboration of the researchers to analyze the data could have positive effects on health care resource planning (Seyed, M. A. et al, 2020).

This could be achieved by extracting data from different spots such as the media, social media networks used by service users and their relatives. However, this approach has limitations as it is dependent on the zeitgeist and the subjective views and biases of those using social media. Another approach would be the close observation and description of service users' needs in hospitals and outpatient contexts (Moritz, E. W. and Frank, U., 2020).

The data acquired by Data Mining can likewise be used for publicizing, extortion recognition and logical discoveries and so on. Data mining is otherwise known as information disclosure, information extraction, investigation of data/designs, data collecting, and so forth (Uzama, T. et al, 2021).

In this research, we analyze the collected data about healthy diet during this COVID-19 for top five African countries that are nowadays struggling in fighting against this pandemic by taking serious and appropriate measures, those countries are Rwanda, Uganda, Namibia, Tunisia, and Zambia as of (Peter, 2021); and we are using deep analysis approach to visualize the insights from data by responding the question "What do we know about non-pharmaceutical interventions in fighting with coronavirus pandemic?".

METHODOLOGY

To carry out this work and achieve its goal, which is to respond on the given question "What do we know about non-pharmaceutical interventions in fighting with coronavirus pandemic?" through data visualization.

Research approach: I am using an analytical-based research based on visualizing collected data in quantitative analyses approach.

Data collection: Data to be used in this study are secondary data collected from https://www.kaggle.com/mariaren/covid19-healthy-diet-dataset, supported by literatures data from existing textbooks, research papers, and thesis.

Data description: In the past year, we have witnessed doctors, nurses, social and thousands of medical workers putting their lives on the frontline to save patients who are infected by COVID-19. All asking ourselves what can we do to protect our loved ones, those who fought for us, and ourselves from this pandemic? These questions and many others, all recount back to the CORD-19 and the answer should be we need to protect ourselves and ours by become accustomed to a healthy diet.

This dataset, combined data of different types of food, world population obesity and undernourished rate, and global COVID-19 cases count from around the world in a period scale between August 2020 and January 2021; in order to learn more about how a healthy eating style could help combat the Corona Virus; in each of the files for data collected it has calculated fat quantity, energy intake (kcal), food supply quantity (kg), and protein for different categories of food (all calculated as percentage of total intake amount); also, added on the obesity and undernourished rate (also in percentage) for comparison.

Data visualization: Visualization is a vital component of the analytics project, but the information presented without context loses some of its insight; insights may take from several parts of a set or combine diverse sources into an actionable understanding. However, without the ability to quickly and logically order it, conveying the information is a less direct process. In this research data is being presented on histograms using Matplotlib, a Python library.

FNDINGS AND DISCUSSION

Findings

As introduced above, we are analyzing data with purpose of understanding other influences known as non-pharmaceutical interventions in fighting with coronavirus pandemic; data collected assess the fat quantity, energy intake (kcal), food supply quantity (kg), and protein for different categories of food (all calculated as percentage of total intake amount) about these food categories: Alcoholic Beverages, Animal Products, Animal fats, Cereals - Excluding Beer, Eggs Fish, Seafood Fruits-Excluding wine, Meat, Milk-Excluding Butter, Vegetal Products, and Vegetable Oils, then we collected data on Deaths, Recovered, and Active ratios considering Rwanda, Uganda, Namibia, Tunisia, and Zambia.
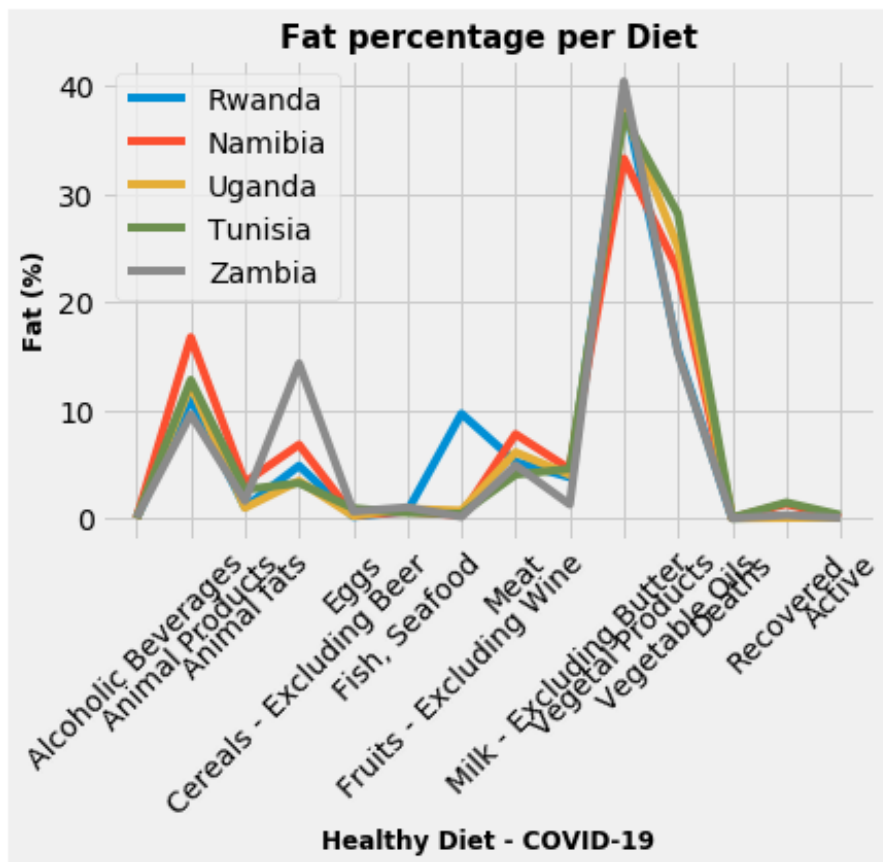
Figure 1 - Fat percentage per food quantity

Visualizing the above data presented using Fig. 1, it is shown that mainly for those countries, food consumed in all four aspects, trends are moving together and differences are occurring on some points for Rwanda and Zambia. Regarding to the food category, in all countries Alcoholic beverages are less consumed in these days, whereas Vegetable products are the most consumed food to supply highest percentage of fat in these days with 38.90%, 33.26%, 37.69%, 37.21%, and 40.39% respectively for Rwanda, Namibia, Uganda, Tunisia, and Zambia. In average, Namibians are the most to consume high fat percentage of 6.99%.
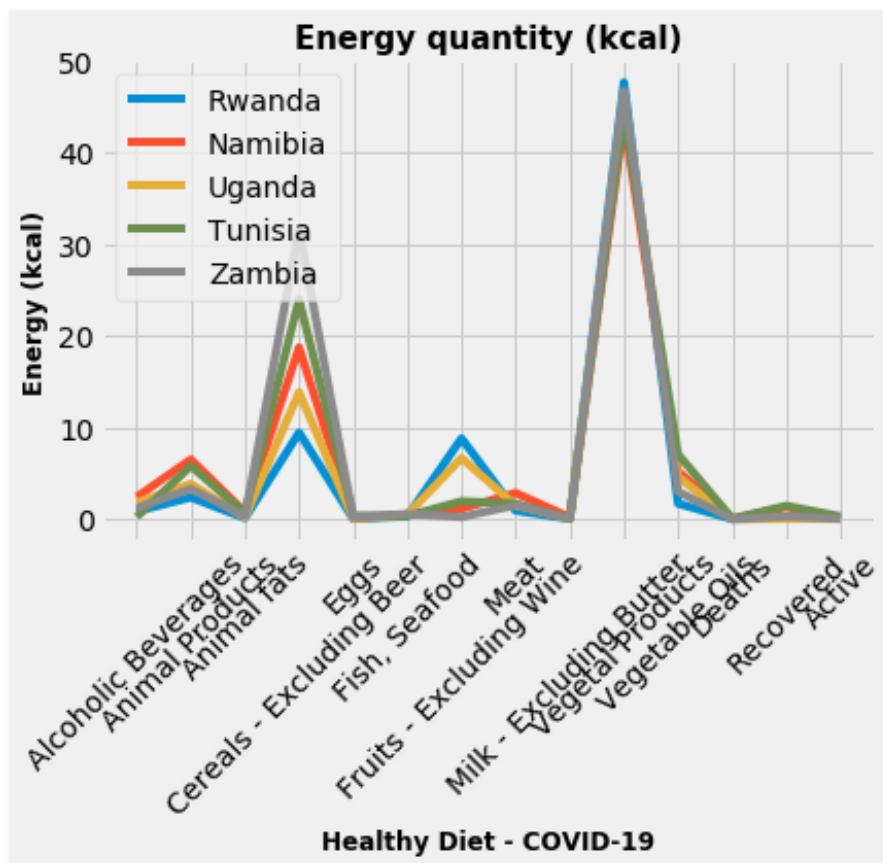
Figure 2 - Energy supply per food quantity (Kcal)

From Fig. 2, the energy acquired from a considered quantity of food indicate that as of the fat percentage, calories of energy for Vegetable Products are the highest comparatively to other food categories; also Cereals (Excluding Beer) are other food category to supply a higher calories of energy for Tunisia and Zambia with 23.95kcal and 31.55kcal respectively. Additionally, other foods like Fruits, Animal Products, Eggs, Fish, and Meat were examined and visualized whereby for Rwanda the averages for each are respectively 8.80kcal, 2.37kcal, 0.02kcal, and 0.92kcal. Reading the dataset, you see that consumed energy from foods is highly acquired from Vegetables with a total of 228.18kcal, and the least one is from milk with 0.33kcal.
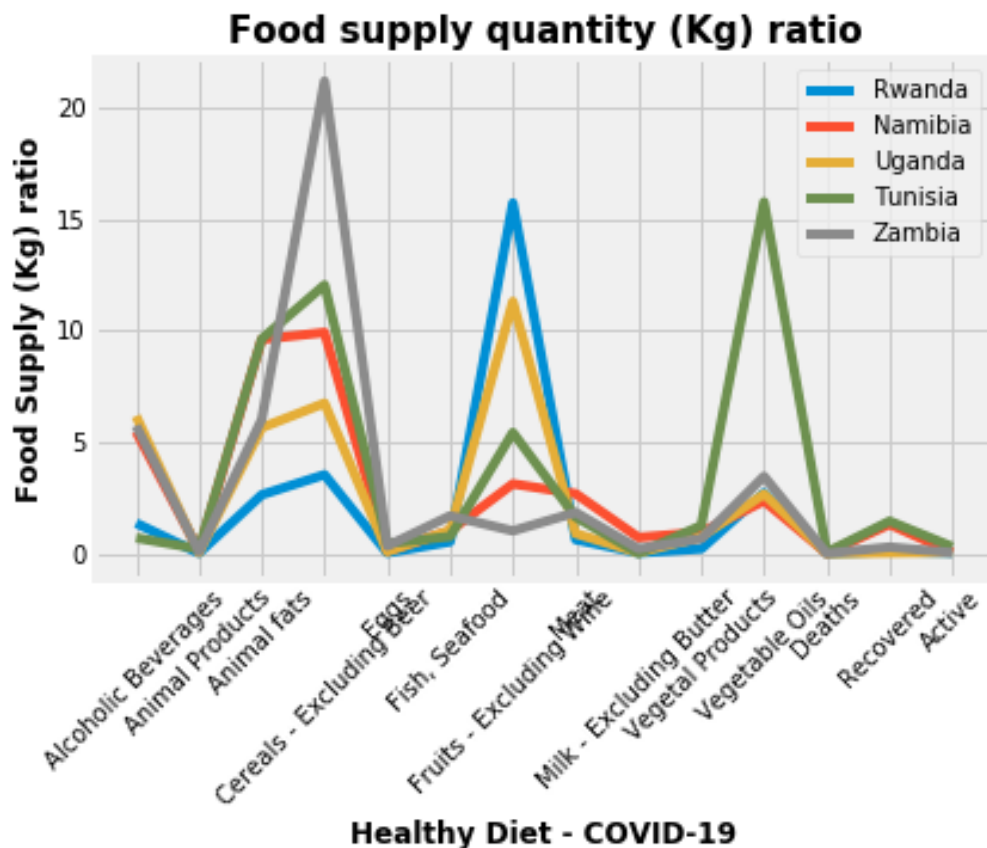
Figure 3 - Food supply quantity (Kg)

The above Fig. 3 shows the image of quantity of food ratio in Kilograms considering foods categories described above; Cereals are the most consumed foods in Zambia with 21.19 kg, in case of Rwanda fruits are the most consumed in quantity at a ratio of 15.72 kg, for Tunisians are consuming Vegetable oils with a ratio of 15.77 kg. In overview you see that the Milk is one of the foods categories which is least in consumptions with a quantity ratio of 0.20 kg, with 1.53 kg for Meat.
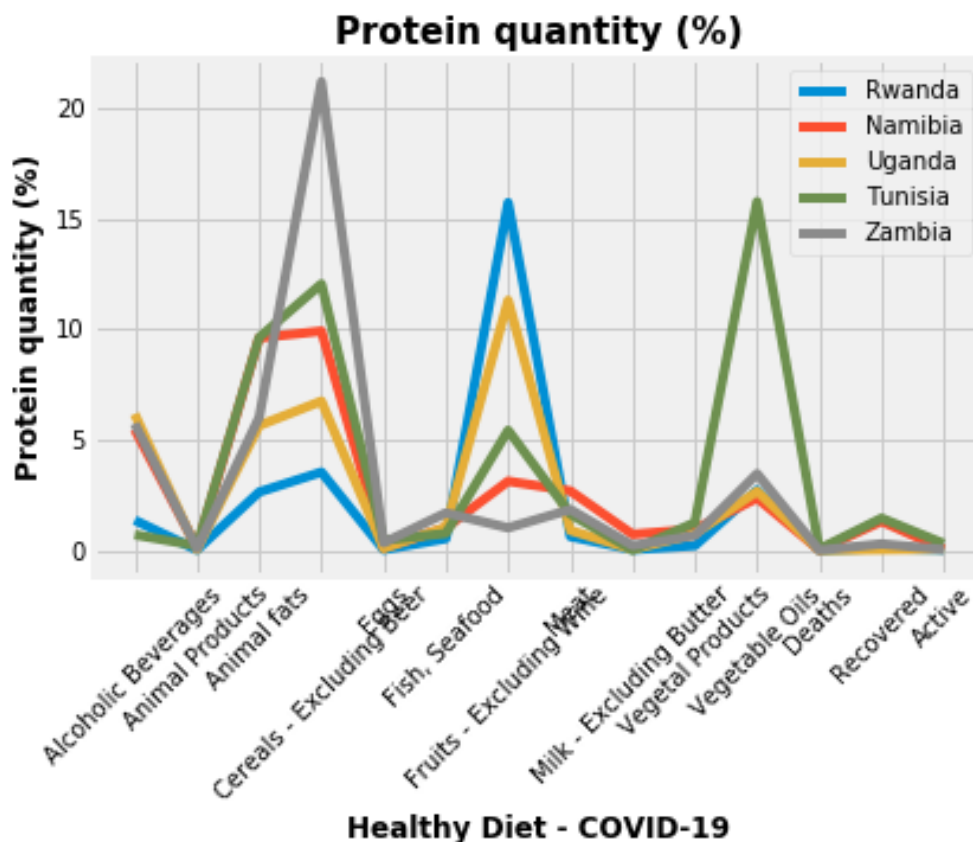
Figure 4 - Protein ratio (%)

This study aims at examining the other side, non-pharmaceutical interventions in fighting with coronavirus pandemic; in this Fig. 4, data shows that trends are likely to move such that of quantity in kilograms; as Zambians are the most consumers of cereals what moves with protein ratio for cereals 28.51% of protein, Namibians that offer highest ratio of protein in animal products and meat with 18.31% and 9.61%; for Tunisians they consumed their proteins from animal products and cereals successively with 14.24% and 25.14% of taken food quantity.

Countries like Rwanda and Uganda are generally not distinguished in such food category like the others studies in this study; regarding the ratio of protein per quantity of foods consumed, you see that Ugandans mainly consume their proteins via animal products and cereals at 10.39% and 13.78% respectively.

Discussions

Just to remind that this study main objectives is to study other non-pharmaceutical inputs that can be taken into accounts as measures of fighting against COVID-19, data about healthy diet for five African countries were collected and analyzed to see how are insights about data then interpret them; it was shown that generally in this period foods such as Cereals, Vegetables and

Meats are the most consumed in these considered countries, this is justified by the averages for these foods categories in fat (%), energy supplied (kcal), and protein (%f) at a ratio of 32.6%, 97.52 kcal, and 96.20% respectively for Cereals; when visualizing the calculated values for Vegetables, the ratio of fat 28.02%, 228.18 kcal of supplied energy, and 0.08% of protein. About Meats you will see that there are 28.02% of fat, 8.47 kcal of energy, and 25.60% of protein acquired form a considered quantity of foods.

Also it is shown by the results that in general the most consumed foods where Rwandans are taking highest rate of fat are Fruits, Cereals, and Vegetables at 9.67%, 38.90%, and 4.85%; while for Ugandans they take proteins mainly in Cereals, and Animal products at a ratio of 13.78% and 10.39%.

CONCLUSION

When concluding, some remarked points of discussions are that during the pandemic population from various countries are taking healthy diet to strengthen their body and help themselves to fight against COVID-19. As of (Roger, 2021), the USDA Center for Nutrition Policy and Promotion recommends a very simple daily diet intake guideline: 30% grains, 40% vegetables, 10% fruits, and 20% protein, but are we really eating in the healthy eating style recommended by these food divisions and balances.

By trying to respond other question of this study, our results indicate that in all countries Alcoholic beverages are less consumed in these days, whereas Vegetable products are the most consumed food to supply highest percentage of fat in these days with 38.90%, 33.26%, 37.69%, 37.21%, and 40.39% respectively for Rwanda, Namibia, Uganda, Tunisia, and Zambia. also Cereals (Excluding Beer) are other food category to supply a higher calories of energy for Tunisia and Zambia with 23.95kcal and 31.55kcal respectively. In overview you see that the Milk is one of the foods categories which is least in consumption with a quantity ratio of 0.20 kg, with 1.53 kg for Meat whereby Rwandans are taking highest rate of fat are Fruits, Cereals, and Vegetables at 9.67%, 38.90%, and 4.85%; while for Ugandans they take proteins mainly in Cereals, and Animal products at a ratio of 13.78% and 10.39%.

REFERENCES

Abdelrahman, E. E. E. et al. (2020). Data Analytics for Predicting COVID-19 Cases in Top Affected Countries: Observations and Recommendations. *International Journal of Environmental Research and Public Health*, 1-25.

FAO. (2020). *Maintaining a healthy diet during the COVID-19 pandemic.* Rome: FAO.

Faseeha, A. and Sadia, M. (2020). How Nutrition can help to fight against COVID-19 Pandemic. *Pakistan Journal of Medical Sciences*, 21-23.

Moritz, E. W. and Frank, U. (2020). Research on COVID-19 and mental health: Data mining reveals blind spots. *Acta Psychiatrica Scandinavica*, 1-8.

Peter, M. (2021, July 09). Coronavirus in Africa: Concern grows over third wave of infections . *BBC News*, pp. 1-2.

Qiong, J. et al. (2020). Big Data Analytics in the Fight against Major Public Health Incidents (Including COVID-19): A Conceptual Framework. *International Journal of Environmental Research and Public Health*, 1-20.

Roger, W. (2021, March). *"Health requires healthy food.".* Retrieved from Kaggle: https://www.kaggle.com/mariaren/covid19-healthy-diet-dataset

Seyed, M. A. et al. (2020). Predicting COVID-19 Incidence Through Analysis of Google Trends Data in Iran: Data Mining and Deep Learning Pilot Study. *JMIR Public Health Surveill*, 17-28.

Shams, M. Y. et al. (2021). Impact of COVID-19 Pandemic on Diet Prediction and Patient Health Based on Support Vector Machine. *Advanced Machine Learning Technologies and Applications, Proceedings of AMLTA*, 64 - 76.

Uzama, T. et al. (2021). Analyzing COVID-19 Dataset through Data Mining Tool "Orange". *2021 2nd International Conference on Computation, Automation and Knowledge Management (ICCAKM)* (pp. 26-40). Dubai, United Arab Emirates: IEEE.

BIODATA

Bob NTAKIRUTIMANA is an Assistant IT Head of ICT and Lab Technician at University of Lay Adventists of Kigali, Rwamagana Campus; and Student at Master of Science in Information Technology, University of Lay Adventists of Kigali. Email: bob.ntakirutimana@unila.ac.rw

Dr. Papias NIYIGENA, PhD; Lecturer at University of Lay Adventists of Kigali.