# Kenko: Individual’s Health Consumption Prediction Using Machine Learning

Prajakta Gurav, Devesh Ramesh
*Mumbai, India.*

**guravprajakta2001@gmail.com, rdeveshr@gmail.com**

***Abstract*— what we eat makes our well-being, be it physically or mentally. The type of food we consume and the food plan we follow is essential. Consuming regular amount of outside/fast food affects the nutritional intake of our body and degrades its functionality resulting in negative outcomes. Excess consumption of junk food leads to constipation, depression, mood swings and food addiction to name a few. Machine Learning (ML) is a branch of artificial intelligence (AI) which uses data and algorithms to imitate the models and eventually improves its accuracy. ML for prediction analyses vast amounts of data and identifies the patterns, relationships, behaviors of models more accurately as compared to humans making it more demanding in the healthcare industry. In this project, ML model is used for predicting how frequent consumption of outside food has negative outcomes on individuals. Initially, a survey is conducted between April 2023 and May 2023 where in, real time data is collected from 150+ various individuals in the age group 18-22 on the factors representing how they consume food items on a daily basis.**

***Keywords*— *Machine learning (ML), Logistic Regression, Support Vector Machine (SVM).***

## I. INTRODUCTION

In today’s fast changing world consuming quality food has become an asset. Due to busy schedules of individuals day to day life, people have accepted the habit of eating any kind of food available at their nearest. The food grown in today’s world is researched to have low nutritional intake due to climatic changes, global warming, greenhouse effects etc. Hence, individuals and the society has to adapt to a lifestyle of fast food availability. But the aftermath of this conventional food intake is affecting the health of individuals on a major extend. Fast food consumption has adverse effects on the health of individuals. Individuals consume food as it is convenient to the taste buds and also is easily available on streets and shops nearby their commute. It therefore becomes easy for them to associate the food with their daily schedule as it is available nearby their work commute. Consuming food which is available from shops and street food can result in health issues such as addiction to that particular food item, feeling less energy after consumption, constipation after few meals, affected mood swings and many more health effects.

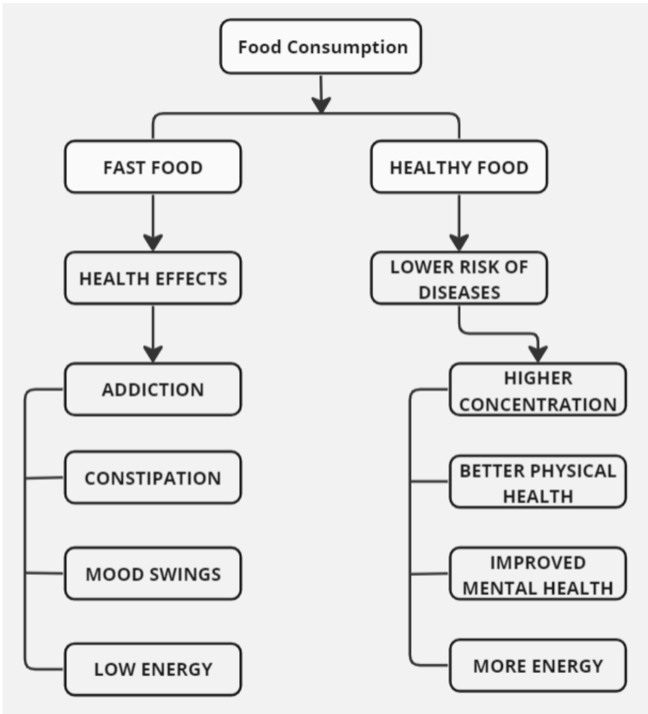


**Fig. 1: Flow Diagram describing the overview of Model.**

Fig1. describes the overall generalization of the data collected. As the consumption of conventional food is increasing, the quality of food intake is decreasing day by day, resulting in weaker health in humans.

Constant intake of food can result in addiction which is due to the individuals consuming the food intake regularly. The reason found from the survey indicated that most of the individuals accepted street food as it was available at their commutes, the food being eaten under peer pressure or living alone.

In [1], AutoML is used for RUL prediction where AutoRUL is the only state of the art model able to generate at least competitive results on a wide variety of datasets. It is able to compete with complex solutions for the domain knowledge. [2] Uses machine learning approach to evaluate live freshness of food. The model achieves high performance for classifying the freshness level of meat. The model is effective in reducing the risks of selling rotten/defective products entailing serious health-based consequences. Paper [3] summarizes a framework for identifying biases in the data and model and mitigating them. The model provides a comprehensive framework for developing accurate predictive models. Paper [4] provides a view on the work done to improve interpretability and accessibility of machine learning in the context of global issues while also being relevant to recent developing countries. This paper uses a data-centric AI system consisting of a model and a learning algorithm. Paper [5] describes the problem of obesity in children and adults using ML datasets for prediction and analysis. The obesity results in chronic diseases. [6] describes how regression models depict variations in wellbeing and their analysis on its respondents. It also tells that tree-based ML perform much better at predicting wellbeing than more conventional linear models in the proposed project. Also, these machine learning approaches validate the previous human guided search for the determinants of its wellbeing. [7], frames the ethics of ML in health care through the lens of social justice. Ongoing efforts and outline challenges in a proposed pipeline of ethical ML are described in health, ranging from problem selection to post-deployment considerations. Paper [8] describes the recent achievements of ML in healthcare specifically deep learning and its digital transformations necessary considerations are outlined from the perspective of current gaps in research, as well as from the lived experiences of health care professionals in resource-limited settings. Authors of [9] proposes a framework, named as 'ML Health', for tracking potential drops in the predictive performance of ML models in the absence of labels as new challenges are uncovered such as monitoring and management of real-time prediction quality of a model in the absence of labels required for deployment. Lastly for [10], in order to reap full benefits of ML in healthcare, we need to realize the limitations of existing datasets and appreciate a system's approach to model the complex healthcare landscape which are associated with datasets and its outcomes.

## II. SYSTEM MODEL

Fig. 2. describes the model evaluation on the basis of the flow of data for machine learning classification. In the initial phase, data is collected on the basis of a survey where in individuals aged 18-23 are asked about their experience associated with conventional food intake in their daily lives. Also, the creation of the dataset has indicated more details of the features. Further, the second phase describes the implementation of Machine learning algorithms namely supervised machine learning for building the model. Here, two algorithms are used which is Logistic Regression and Support Vector Machine. Their accuracies are compared for model understanding and evaluation. Third phases describes the creation of dashboard for data analysis and data mining features with respect to key performance indicators (KPIs). In the final phase the dataset is queried using Structured Query Language (SQL) as an alternate method for predicting insights of the dataset for making data driven decisions.

The Fig (2). Below describes the flow of ML algorithms for model evaluation and Fig (3) depicts necessary features of the dataset.

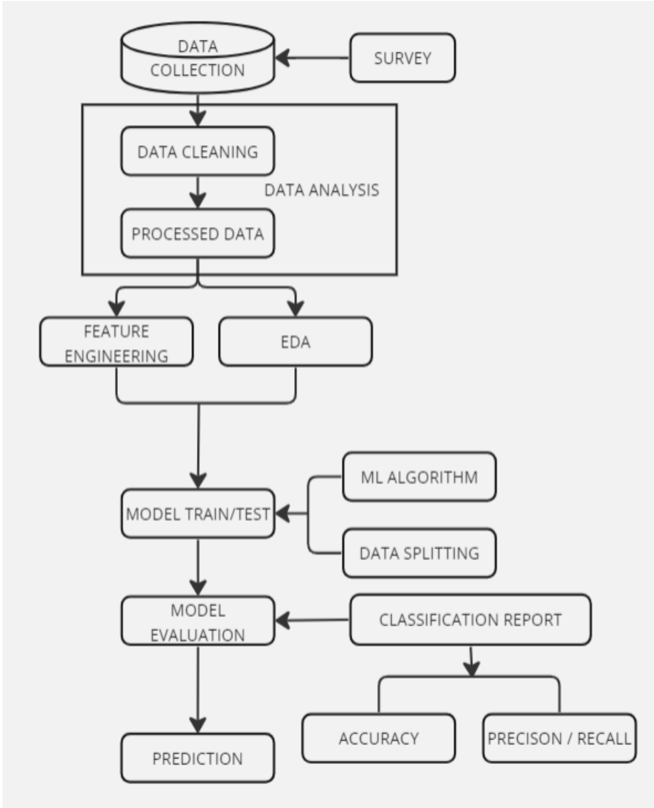


**Fig. 2: Proposed Methodology for Machine Learning Model Prediction.**

## I. RESULTS AND DISCUSSIONS

In this section, we present the performance of three ML models employed for health prediction. For Logistic Regression model the data is evaluated using various plots and relationship between the features is evaluated using confusion matrix. Fig. 5 illustrates the confusion matrix of total data collected. In case of SVMs, kernels are compared for better understanding of performance. Table 1 describes the accuracy of machine learning models which performed more efficient. SVM has performed better than logistic regression in terms of model classification and performance.

| Model | Accuracy (%) |
|---|---|
| Logistic Regression | 67.741 |
| SVM | 69.354 |

**Table 1: model accuracy**

Fig. 5 describes the overall relationship between all the features of the dataset where in the variables are related with one another as a relation between them. The features represent the different descriptions of a confusion matrix of true positive, true negative, false positive and false negative with 0 and 1 values.

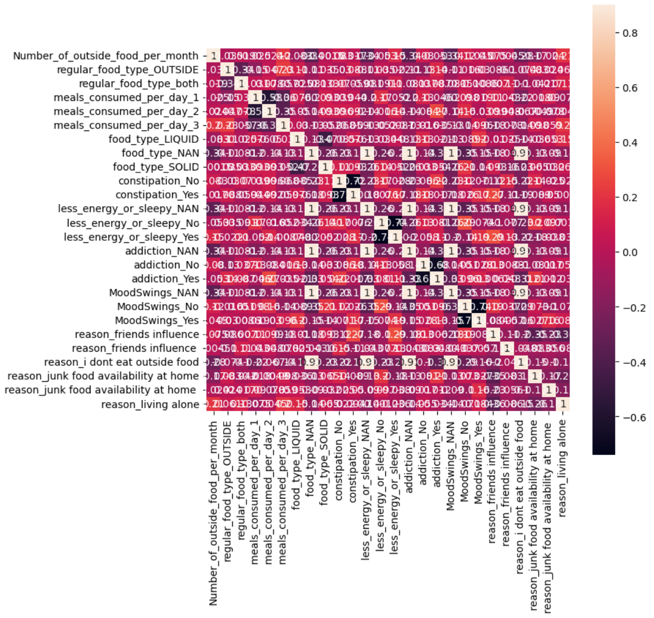


**Fig. 5: Confusion Matrix**

Fig.6 describes the intake of individual's regular convenient food consumption as per the density graph. Also, the below Table 2. depicts the evaluation of machine learning model for the algorithm logistic regression where the accuracy of the model was found to be 67.741% and the Roc Auc score was found to be 56.181%.

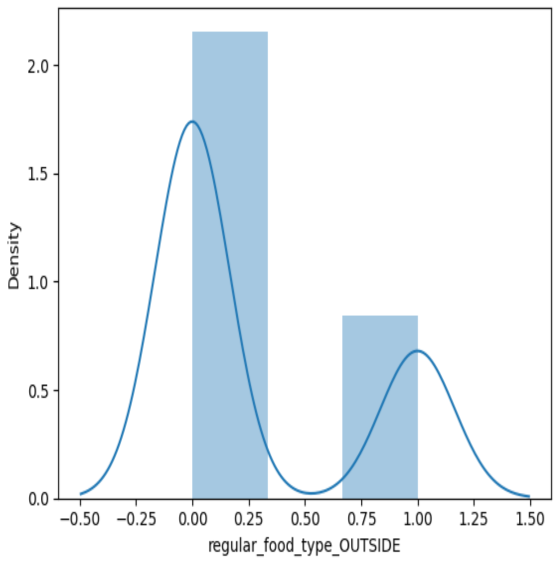


**Fig. 6: Consumption of regular food type with respect to density**

AUC-ROC stands for 'Area under the Curve' of the 'Receiving Operating Characteristics' of the curve. An AUC-ROC curve is an evaluation metric which is often used in a classification machine learning algorithm to visualize its performance. The AUC-ROC score represents how efficient the model performs.

**Table 2: Logistic regression model evaluation**

| Logistic Regression parameters | Percentage (%) |
|---|---|
| Accuracy | 67.741 |
| Roc Auc score | 56.181 |

Below given Table 3. Depicts the classification report for logistic regression model. It shows that the precision and recall for '0' is 73% and 86% respectively whereas the same for '1' is 45% and 26%. The F1 score interpreted is 68% for accuracy with 56% for macro average and 65% for weighted average.

**Table 3. Classification Report for Logistic Regression**

| Models Logistic Regression | precision | recall | F1-score | support |
|---|---|---|---|---|
| 0 | 0.73 | 0.86 | 0.79 | 43 |
| 1 | 0.45 | 0.26 | 0.33 | 19 |
| accuracy | - | - | 0.68 | 62 |
| Macro avg | 0.59 | 0.56 | 0.56 | 62 |
| Weighted avg | 0.64 | 0.68 | 0.65 | 62 |

Evaluation of SVM algorithm describes that RBF kernel has the highest accuracy with precision and recall of 0 respectively. Whereas, the precision for polynomial and sigmoid kernels is 33.33% with sigmoid having higher recall of 15.7%.

**Table 4. Evaluation for SVM**

| SVM kernels | parameters | Percentage (%) |
|---|---|---|
| polynomial | accuracy | 66.125 |
| | Precision | 33.333 |
| | recall | 10.526 |
| Radial basis function(RBF) | accuracy | 69.354 |
| | Precision | 0.0 |
| | recall | 0.0 |
| Sigmoid | accuracy | 64.516 |
| | Precision | 33.333 |
| | recall | 15.789 |

Classification report of SVM describes precision and recall for '0' as 76% and 86% respectively and 33% and 16% for '1'. The F1 score as 49% for macro average and 65% for weighted average.

**Table 5. Classification Report for SVM**

| Models Logistic Regression | precision | recall | F1-score | support |
|---|---|---|---|---|
| 0 | 0.70 | 0.86 | 0.77 | 43 |
| 1 | 0.33 | 0.16 | 0.21 | 19 |
| accuracy | - | - | 0.65 | 62 |
| Macro avg | 0.52 | 0.51 | 0.49 | 62 |
| Weighted avg | 0.59 | 0.65 | 0.65 | 62 |

Given below Fig 7. Describes the different types of food consumed by individuals in their daily schedules. It depicts home-made type to be the most consumed followed by outside in second place. It is also seen that the consumption of both types is also not less with slightly more than 50%.

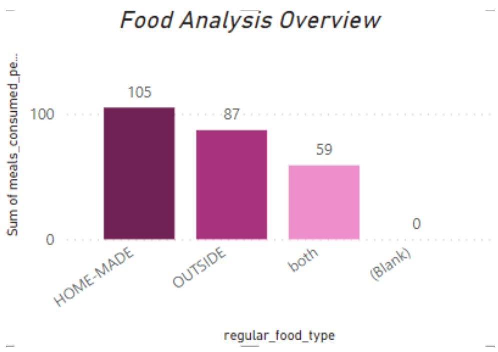


**Fig 7.**

Fig8. Describes the addiction ratios of individual's food intake wherein they get addicted to fast food due to regular intake. It shows almost 66.53% individual's addiction to fast food after few meals followed by 10.76% for may or may not type. The individuals having no addiction is almost 21.91% as they do not consume fast food.

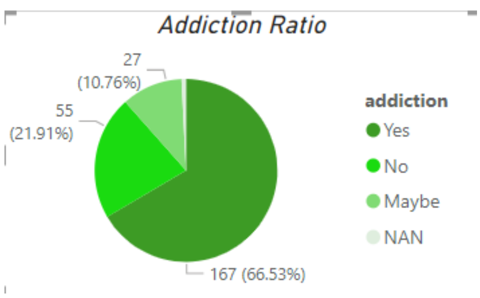


**Fig 8.**

Fig9. Describes the constipation ratios of individual's food intake. It shows almost 31.8% individual's adverse effects to fast food after few meals followed by 12.0% for may or may not type. The individuals not consuming fast food are almost 54.58% as they represent better physical health.

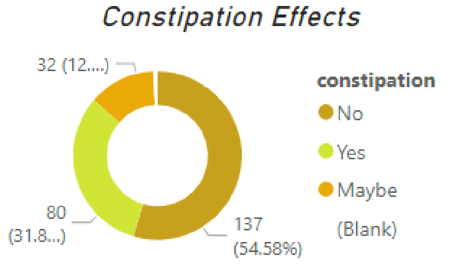


**Fig 9.**

Fig.10. Describes the Mood Swings analysis of individual's food intake. It shows almost 65% individual's result in mood swings due to fast food intake just after few meals followed by 15% for may or may not type. The individuals not affected to mood swings is almost 170% as they do not consume fast food.

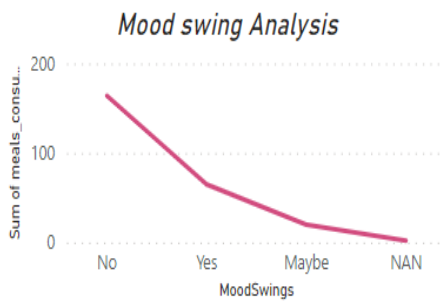


**Fig 10.**

Fig.11. Describes the Energy status of individual's food intake. It shows moderated result in energy status due to fast food intake just after few meals followed by least for may or may not type. The individuals who are not affected to energy and sleepiness their fluctuation is almost minimum as they do not consume fast food.

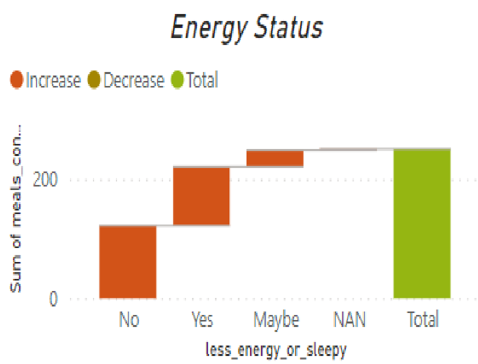


**Fig 11.**

## II. CONCLUSION

In this project, predicting the health of individuals on the intake of fast food was proposed where in dataset was created on the basis of a survey conducted. The results of the model predict that individuals who consume more quantity of fast food regularly are more prone to various health-related issues as compared to individuals who do not consume conventional food more often. Machine Learning models depicted that logistic regression has an accuracy of 67% for prediction for more than 150 individual responses in the dataset. Whereas, for SVM, the Radial Basis Function(RBF) performs better as compared to Polynomial and Sigmoid kernels in terms of accuracy.

Furthermore, individual observation of features depict that majority of individuals consume conventional fast food due to unavoidable reasons such as living alone or under the influence of peers. This paper demonstrates the potential outcomes individuals face due to conventional food intake. It was also found that 66.53% individuals suffer from addiction after few intakes, 31.8% have constipation, almost 66% have mood swings and 25% decrease in energy by feeling of sleepiness or fatigue. Those individuals who do not consume fast food have a healthier lifestyle. Also, they are less prone to health risks and better well-being.

## III. LIMITATIONS

The data collected for algorithm training shows moderated results due to less data acquired. However, the model can predict better accuracy and more improved results for more number of iterations of a larger dataset. This will further enhance the prediction of adverse health effects to people and motivate them to indulge in a healthy lifestyle.

## IV. REFERENCES


[1] Marc-André Zöller, Fabian Mauthe, Peter Zeiler, Marius Lindauer and Marco F. Huber, “Automated Machine Learning for Remaining Useful Life Predictions” in arXiv:2306.12215v1 [cs.LG] 21 Jun 2023.

[2] BhargavSagiraju1, NathanCasanova1, LamIvanChuenChun1 MananLohia1 ToshinoriYoshiyasu1, “Meat Freshness Prediction”, in arXiv:2305.00986v1 [cs.CV] 1 May 2023.

[3] Shaina Raza, “Connecting Fairness in Machine Learning with Public Health Equity”, in arXiv:2304.04761 [cs.LG] 8 Apr 2023.

[4] N Ranasinghe, A Ramanan, S Fernando, PN Hameed, D Herath, T Malepathirana, P Suganthan, M Niranjan and S Halgamuge, “Interpretability and Accessibility of Machine Learning in selected Food Processing, Agriculture and Health Applications”, in arXiv:2211.16699 [cs.LG] 30 Nov 2022.

[5] Amin Gasmi, “Machine Learning and Bioinformatics for Diagnosis Analysis of Obesity Spectrum Disorders”, in arXiv:2208.03139 [q-bio.QM] 5 Aug 2022.

[6] Ekaterina Oparina, Caspar Kaiser, Niccolo` Gentile, Alexandre Tkatchenko, Andrew E. Clark, Jan-Emmanuel De Neve, Conchita D’Ambrosio, “Human Wellbeing and Machine Learning” in arXiv:2206.00574v1 [econ.EM] 1 Jun 2022.

[7] Irene Y. Chen, Emma Pierson, Sherri Rose, Shalmali Joshi, Kadija Ferryman, and Marzyeh Ghassemi, “Ethical Machine Learning in Health Care” in arXiv:2009.10576v3 [cs.CY] 8 Oct 2020.

[8] Varoon Mathur, Saptarshi Purkayastha, Saptarshi Purkayastha, “Artificial Intelligence For Global Health: Learning From A Decade Of Digital Transformation In Health Care”, in arXiv:2005.12378v2 [cs.CY] 27 May 2020.

[9] SindhuGhanta, SriramSubramanian, LiorKhermosh, Swaminathan Sundararaman, Harshil Shah, YakovGoldberg, DrewRoselli, Nisha Talagala, “Ml Health: Fitness Tracking For Production Models” in arXiv:1902.02808v1 [cs.LG] 7 Feb 2019.

[10] Awais Ashfaq, Slawomir Nowaczyk, “Machine learning in healthcare - a system’s perspective” in arXiv:1909.07370v2 [cs.CY] 19 Jan 2020.