Minimizing Food Wastages and Predicting Raw Materials required for Cloud Kitchens using Machine Learning

Corresponding author: Jawed Hawari

Jawed Hawari
Department of Computer Science and Engineering
Amrita School of Computing, Bangalore
Amrita Vishwa Vidyapeetham, India
BL.EN.U4CSE21073@bl.students.amrita.edu

Nikunj Jain
Department of Computer Science and Engineering
Amrita School of Computing, Bangalore
Amrita Vishwa Vidyapeetham, India
BL.EN.U4CSE21138@bl.students.amrita.edu

Priyanshu Jha
Department of Computer Science and Engineering
Amrita School of Computing, Bangalore
Amrita Vishwa Vidyapeetham, India
BL.EN.U4CSE21165@bl.students.amrita.edu

Abstract—The increase in frequency and severity of extreme weather events sometimes causes a food wastage in the kitchen which works on cloud platform. Our model demonstrates, how we can reduce the wastage of food with help of machine learning applying on weather data and food data.

Keywords—Machine learning, Clustering Analysis, Segmentation

I. INTRODUCTION

According to a recent report from the Global Hunger Index (GHI) [1], India finds itself ranked 100th out of 119 countries facing severe hunger challenges. What's even more concerning is that this places us behind some of our neighboring countries like China, Nepal, and Sri Lanka. It's a stark reminder of the hunger crisis we're grappling with, as approximately 190 million people in India go to bed hungry every night. Shockingly, this includes 35% of women and 28% of men.

One major contributing factor to these distressing statistics is the significant food wastage that occurs in our country. Every year, tons of food go to waste from various sources such as organizations, restaurants, and many other food providers. This wastage is a tragedy because simply increasing agricultural productivity cannot offset the loss of food. Challenges like climate change, dwindling land availability, and water scarcity make it incredibly tough to enhance food production

II LITERATURE SURVEY

This section entails the literature of work done on food management

2.0 **Food Waste Management Using Machine Learning**

One of the most important causes of this problem is food waste, a recurring theme in industries such as institutions and restaurants. The consequences of food waste are compounded by the fact that increased agricultural production alone cannot compensate for this loss India faces severe challenges such as climate change, soil availability and water scarcity is rare, which makes increasing food production an increasing workload.

Current literature emphasizes the need for alternative approaches to hunger management in India. One promising approach to research is the use of supervised learning processes, which can provide valuable insights and solutions. These algorithms have the potential to contribute significantly to efforts aimed at reducing hunger and increasing food security in the country. This literature review presents a way to evaluate the proposed approach in the context of solving the Indian hunger problem.

2.1 **Problem formulation**

- The main issue we face comes from its root: food waste. According to many research findings, dividing the total global food production into three equal parts reveals a sad fact: One-third of this valuable resource is wasted Food a pollution has always been a concern since time immemorial, evading our adequate solutions even in our modern age of advanced technology and knowledge.

The main objective of our research project is to solve the widespread problem of food waste by harnessing the potential of machine learning. Our ambition is to lead new machine learning paradigms to alleviate this issue. However, as a first step toward this nobler goal, we have developed a practical way to engage: implementing an machine learning model in our dormitories in the 19th century.

Our strategic plan includes developing and implementing a new ML model specifically designed to manage food waste in cloud kitchen.

A key feature of its prediction function is that the **system** actively recommends the best amount of raw materials that the restaurant should buy in the near future. This recommendation process is based on aligning a machine-learning model with the business priorities of the restaurant. If the restaurant agrees with the plan's recommendations, they are free to authorize the order. The application then sets up a process to easily deliver this order to restaurants from trusted distributors.

A possible way to drastically reduce food wastages is using weather forecasting which will help us in managing raw material and eventually reduces wastes
We will be using following models for our project:
1.Numerical weather prediction models
2.classification algorithm
3.Regresion algorithms
4.support vector machines
5.deep learning
So we will using these tools for further research and analysis

III DATA DESCRIPTION

The dataset utilized in this experiment is from the Excel file Swiggy dataset which contains food delivery time along with restaurant name, time, location and various other factors. These datasets form the basis of our research, with KNN algorithms and techniques used to extract significant insights in their respective disciplines.

IV METHODOLOGY

In this section, we describe the methodology used for customers and restaurant, adhering to IEEE format guidelines.

1. ***Well separation of Classes in dataset***: Well separation of two classes can be observed by considering their intra class spread and inter class spread value. In my dataset the values of intra class spread and inter class spread are as follow:

    Intra_class_spread_price = 230.9267
    Intra_class_spread_Total_rating = 391.425
    Inter_class_spread = 191.8096

    Here the inter class spread is smaller than the intra class spread of each classes. That means the classes in dataset are fairly well separated.

2. ***Measure of class separation using class centroids***: The distance between two centroids (or Euclidean distance) is good enough measure to test for class separabilty. Larger the Euclidean distance means greater class separation. In our dataset, the value of Euclidean distance is 191.8096, which means that classes are separated.

3. ***Behaviour of kNN classifier based on k value***: kNN classification classifies a data point by looking at the class labels of its k-path nearest neighbours in the training data set. The behaviour of kNN classifier changes with the change of k values.

    i. ***K = 1***: When k is set to a small value (k = 1) the algorithm looks at the nearest neighbors of each data point very carefully and this can make the classifier sensitive to noise and outliers in the data. The decision boundary can be quite irregular and may not be very sensitive to new information that is not observed. This can make a huge difference and over correlation.

    ii. ***K = 5 to K = 20***: It captures more neighbours, reducing noise and peripheral effects. The decision limit is more stable, and the model performance is generally improved on unobserved data. This approach tends to strike a good balance between bias and contrast.

    iii. ***K>20***: As k becomes larger, the decision boundary becomes smoother and also less sensitive to local variations in the data, although this reduces the risk below the threshold of overfitting though, a poor fit could occur if k is too large. The classifier may be too flexible and fail to capture underlying patterns in the data.

Overfitting and Underfitting conditions in kNN

i. ***Overfitting:*** Using a small k value (e.g., k=1) makes the model very sensitive to noise or outliers. It allows data points to be segmented based on individual noisy data points, resulting in over compressed images. Decision boundaries can be highly irregular, resulting in poor generalization to new data.

ii. ***Underfitting***: Using large k values: If you use a very large k value, the decision constraints of the model are very weak. It distributes over the lines of a large number of neighbours and consequently may not be able to capture complex patterns in the data. This can lead to underfitting, where the model has high bias and low variance but performs poorly on both the training and testing data sets.

4. *kNN classifier vs various metrics*: We can state kNN classifier is good classifier or not by analysing the metrics Accuracy, Precision, Recall, F1-Score, etc.

    i. *Accuracy*: Accuracy measures the overall accuracy of your classification. High precision indicates that the classifier has correctly classified a large proportion of the data

    ii. *Precision & Recall*: Precision refers to the accuracy of a good prediction, whereas recall refers to the ability to correctly identify all positive cases in real positive cases.

    iii. *F1-Score* : F1-Score is the harmonic mean of accuracy and recall. It provides a metric that balances accuracy and recall. A good classifier has a high F1-score.

    Values of accuracy, precision, recall, and F1-score for class 'Price'

	precision	recall	f1-score	support
0.0	0.00	0.00	0.00	2
1.0	0.00	0.00	0.00	1
2.0	0.00	0.00	0.00	0
5.0	0.00	0.00	0.00	1
10.0	0.00	0.00	0.00	0
15.0	0.00	0.00	0.00	0
20.0	0.00	0.00	0.00	0
30.0	0.00	0.00	0.00	0
40.0	0.00	0.00	0.00	0
45.0	0.00	0.00	0.00	0
48.0	0.00	0.00	0.00	1
50.0	0.00	0.00	0.00	4
60.0	0.00	0.00	0.00	1

...

	precision	recall	f1-score	support
accuracy			0.16	2604
macro avg	0.03	0.02	0.02	2604
weighted avg	0.16	0.16	0.14	2604

Values of accuracy, precision, and F1-score for class 'Total ratings'

	precision	recall	f1-score	support
20	0.20	0.32	0.25	466
50	0.16	0.17	0.17	300
80	0.57	0.62	0.60	935
100	0.33	0.25	0.29	590
500	0.21	0.07	0.11	178
1000	0.43	0.10	0.16	130
5000	0.00	0.00	0.00	3
10000	0.00	0.00	0.00	2
accuracy			0.37	2604
macro avg	0.24	0.19	0.20	2604
weighted avg	0.37	0.37	0.36	2604

IV CONCLUSION

Our data analysis, data visualization, and preprocessing efforts shed light on important aspects of dealing with complex datasets. We successfully dealt with missing data, standardizing our dataset using the KNN algorithm. These findings provide a solid foundation for data-driven decision-making processes across sectors. It is important to recognize the continued development of analytical methods and the possibility of further research and adaptation in the work. A system designed for customer/restaurant splits effectively incorporates these options.

5. *Regular fit situation in model*: Regular fit of model means the model has predicted right dataset using train model. In our data set the model does not have regular fit because for 'Price' accuracy is extremely low, that means the Price value predicted by model has great difference with its actual value.

6. *Situation of overfit in kNN classifier*: Overfitting in kNN occurs when the model is too sensitive to noise and changes in the training data, resulting in poor generalization performance on new, unseen data. Here are the situation in which overfitting occurs:

    i. *Small value of k*: When you choose a small value of k, such as 1 or 2, the kNN classifier is very sensitive to the individual data points in the training data set and tries to overfit the training data only if very limited he considers the number of nearest neighbours.

    ii. *Noisy data*: Smaller values of k cause the model to include noisy or outlier data points in its decision-making process. These outliers do not represent the true underlying structure of the data and can lead to confusing predictions.

    iii. *Too many features*: Too many features make the model overfit. In kNN classifier, classifier considers only the closest features. Since there are lot of features in dataset there are it become congested and model captures all those irrelevant features, which are closest to target value.