**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**—Extreme weather events increase in frequency and intensity sometimes resulting in food waste in a kitchen operating on a cloud platform. Our example shows, how we can reduce food waste with the help of machine learning applied to weather data and food data. Keywords—machine learning, cluster analysis, classification   1. INTRODUCTION   According to a recent report from the Global Hunger Index (GHI) [1] , India ranks 100 out of 119 countries facing severe hunger challenges What is more worrying is that this puts us behind some neighboring countries such as after China, Nepal and Sri Lanka. It is a stark reminder of the hunger crisis we are dealing with, as some 190 million people in India go hungry every night. Shockingly, it’s 35% women and 28% men. One of the main reasons behind these dismal statistics is the huge amount of food wasted in our country. Every year, tons of food goes to waste from a variety of sources, including institutions, restaurants, and many other food service providers. This waste is unfortunate because increases in agricultural production alone cannot offset food losses. Challenges such as climate change, land degradation and water scarcity make it very difficult to increase 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, which is a common issue in industries such as institutions and restaurants. The consequences of food waste are compounded by the fact that increased agricultural waste alone cannot compensate for this loss India rarely faces severe challenges such as climate change, land availability there and water scarcity, making increasing food production an increasing task  The main feature of this prediction task is that the system actively recommends the optimal number of raw materials that the restaurant should purchase in the near future This recommendation system is based on the machine learning model and restaurant of business priorities are aligned. If the restaurant agrees with the planning proposal, they are free to authorize the order. The app then sets up a system for trusted distributors to make it easier for them to deliver the order to restaurants  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.   1. ***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. 2. ***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. 3. ***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. 4. ***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. 5. ***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.   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, recall, 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   1. *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. 2. *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: 3. *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. 4. *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. 5. *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. | 2.1 Problem formulation-The main issue we face comes from its root: food waste. Dividing all global food production into three equal parts according to the findings of many studies reveals a sad truth: one-third of this valuable resource is wasted Food waste has been a cause for concern always since antiquity, and evades our proper solutions even in our modern times The main objective of our research project is to solve the prevalent problem of food waste by harnessing the power of machine learning. Our goal is to pioneer new machine learning paradigms to alleviate this issue. But as a first step toward this lofty goal, we developed a practical way to get involved: implementing a machine learning prototype in our dormitories at century in the 19th edition.Our strategic plan includes developing and implementing a new ML model specifically designed to manage food waste in cloud kitchen. The main feature of this prediction task is that the system actively recommends the optimal number of raw materials that the restaurant should purchase in the near future This recommendation system is based on the machine learning model and restaurant of business priorities are aligned. If the restaurant agrees with the planning proposal, they are free to authorize the order. The app then sets up a system for trusted distributors to make it easier for them to deliver the order to restaurants  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 use these tools in our research for the completion of our projects    **Overfitting and Underfitting conditions in kNN**     1. ***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. 2. ***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. 3. *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.      1. *Accuracy*: Accuracy measures the overall accuracy of your classification. High precision indicates that the classifier has correctly classified a large proportion of the data 2. *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. 3. *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  ...  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. |