Prediction of the pizza brands with different classification methods

Spring Semester, 2022

1. Introduction

As it is known to us, pizza is a kind of food, which consists of many food ingredients and is in the shape of the circle. In fact, pizza also attracts the appetite from many people, and there are many pizza restaurants all over the world, especially in the European countries like Finland.

Since it is a course project in the Machine Learning field, then it is expected to explore how the machine-learning methods could be applied to the realistic problems. Here want to predict the pizza brands based on the given dataset and information, with machine-learning methods. Moreover, this topic enables people to know the brand of pizza they want to buy, if they are provided with the information of pizza, like the size, the price or the type. Then people would know better about the pizza brands, which helps them to decide the brand they intend to purchase the next time.

Regarding the structure of the project report, the clarification on this machine learning problem is presented below in the Section 2 "Problem Formulation", while the main methods for this project will be presented in the Section 3. After these sections, the Section 4 and Section 5 are expected to show the result and draw the conclusion of the project respectively.

2. Problem Formulation

In terms of the problem of project, we could start to assume that one consumer knows the size, the taste and the price of the pizza, but this person is unaware of the brand of this pizza. Then problem arises: What is the brand of the pizza?

This dataset comes from the online open-source platform Kaggle [1]. The data points of the problem are the record of one pizza, which includes the information of its original company name, the pizza name, the type, the size, as well as the price. The total number of the data numbers is 371. Since the information except the price from the dataset are not in the form of numbers, we could use numerical values to indicate the information. For example, since there are only four brands of the pizza (Domino, Pizza Hut, Godfather, IMO), the numbers from 1 to 4 are used to indicate the brand of pizza respectively. Therefore, it could be a multiclass classification problem, where there are 4 potential categories of outcomes.

2.1 Labels and Features of the data points

This part will introduce the labels and features of the adopted data points in this project briefly.

Label: The brand of the pizza.

Features: The type, the size, and the price of the pizza. (3 features)

3. Methods

4. Results

5. Conclusion

6. References

[1] https://www.kaggle.com/knightbearr/analysis-pizza-price-data-knightbearr/data

Appendix