**EVALUATION OF CUSTOMER SPENDING ON FOOD PRODUCTS USING MACHINE LEARNING**

**Table of Contents**

[Chapter 1: Introduction 2](#_Toc180593944)

[1.1 Context of the research 2](#_Toc180593945)

[1.2 Problem Statement 3](#_Toc180593946)

[1.3 Purpose of this project and its application 3](#_Toc180593947)

[1.4 Research questions 3](#_Toc180593948)

[1.5 Research aim and objectives 4](#_Toc180593949)

[1.5.1 Aim 4](#_Toc180593950)

[1.5.2 Objectives 4](#_Toc180593951)

[1.6 Research Novelty 4](#_Toc180593952)

[Chapter 2: Background 5](#_Toc180593953)

[2.1 Introduction to the literature 5](#_Toc180593954)

[2.2 Selection of the Papers 5](#_Toc180593955)

[2.3 Critical analysis of key papers 7](#_Toc180593956)

[2.4 Comparison of the papers 10](#_Toc180593957)

[2.5 Conceptual Framework 10](#_Toc180593958)

[2.6 Literature Gaps 11](#_Toc180593959)

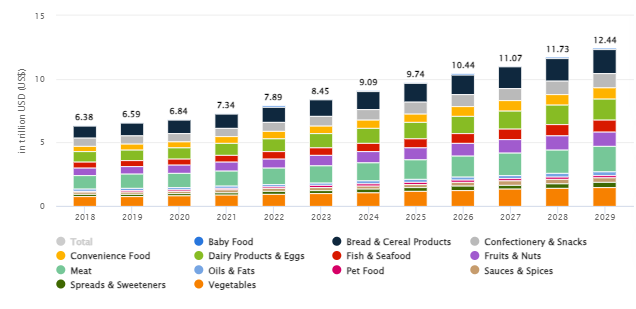
[2.7 Summary 11](#_Toc180593960)

[References 12](#_Toc180593961)

# Chapter 1: Introduction

## 1.1 Context of the research

The term ‘Customer Spending’ can be defined as the total amount of money that consumers spend on goods and services, which primarily includes two components such as induced consumption and autonomous consumption. According to Cai (2023); Daroch, Nagrath and Gupta (2021), the factors influencing consumer spending include economic factors such as income, tax rates, consumer confidence, product pricing, market conditions and customer preferences. Therefore, changes in customer product pricing and income levels can lead to variations in customer spending on food products, as it is linked with the change in market conditions. Global customer expenditure on foods and beverages has reached $9.09 trillion in 2024 (which is approximately 10.7% of total customer expenditure) (Cognitive Market Research, 2023). Additionally, with the increasing population and income, market revenue in the food industry (globally) is expected to reach $12.44 trillion by 2030 (Statista, 2024) (***Refer to Figure 1***). Therefore, within this increasing market trend, the identification of spending patterns enables personalised marketing and ensures efficient resource allocation for better profitability and customer retention.

******

***Figure 1: Market revenue in the food sector across the world from 2018 to 2030***

(Source: Statista, 2024)

## 1.2 Problem Statement

Evaluation of customer spending behaviour within the highly competitive, growing market is crucial for retail food businesses to optimise their operations and increase operational sustainability. Traditional methods (such as Recency, Frequency and Monetary (RFM) analysis and Basket analysis) of analysing customer spending are often limited due to their lack of ability in uncovering complex relationships between customer spending and external attributes like spending preferences (Heldt, Silveira and Luce, 2019; Christy *et al.*, 2018). With the growing availability of large datasets capturing demographic information and purchasing habits, there is an opportunity to leverage machine learning techniques for more accurate and insightful predictions of customer spending on food products.

## 1.3 Purpose of this project and its application

The purpose of this project is to leverage Machine Learning (ML) techniques (such as Clustering and Classification) to predict customer spending on food products based on customer attributes and spending behaviour. This can enable businesses to make data-driven decisions related to consumer spending on food products and their buying behaviour, allowing food businesses within the retail landscape to optimise resource allocation and streamline their productions. The application of this research can be extended to fields like retail management, customer spending prediction and supply chain optimisation as this study has the capability to predict customer spending on food products and their spending behaviour, which can be beneficial for businesses to manage their material resources and optimise their supply chain networks.

## 1.4 Research questions

* What factors influence customer spending on food products and their spending behaviour?
* What is the significance of statistical analysis and Machine Learning techniques in the prediction of customer spending?
* What is the implication of ML algorithms (Clustering and Classification) in predicting customer spending on food products based on historical spending and demographic attributes?

## 1.5 Research aim and objectives

### 1.5.1 Aim

The aim of this study is to evaluate customer spending on food products using Machine Learning (ML) techniques.

### 1.5.2 Objectives

* To evaluate different factors influencing customer spending on food products
* To assess the significance of Machine Learning (ML) techniques for evaluating customer spending on food products
* To develop ML models (Clustering and Classification) for predicting customer spending on food products based on historical spending data and demographic attributes
* To recommend effective strategies to retail food businesses for enhancing customer spending on food products and implying effective market strategies

## 1.6 Research Novelty

The novelty of this study lies in the integration of advanced ML techniques with customer spending prediction models tailored explicitly to the retail food sector. Unlike conventional methods (RFM analysis and Basket analysis) followed in the majority of past studies that are based on linear relationships and simplistic assumptions, this project intends to explore non-linear, high-dimensional relationships in large datasets to unleash underlying patterns in spending behaviour of customers. Additionally, this study has incorporated diverse features like region as well as customer historical spending, this study provided more accurate and actionable predictions for estimating customer spending on food products. Additionally, the incorporation of clustering technique can allow this study to perform customer segmentation based on purchasing habits and spending.

# Chapter 2: Background

## 2.1 Introduction to the literature

Literature reviews refer to evaluation of methodological choices followed in past studies and their respective findings within the context of a specific research field. Literature review in study is based on evaluation of purpose, methodological choices and findings of five selected peer-reviewed papers related to customer spending on food products and implication of ML algorithms in predicting customer spending and customer behaviour. Prediction of customer spending behaviour based on historical spending data is linked with the subject area of Data Science and Machine Learning that integrates concepts of data distribution, Exploratory Data Analysis and ML modelling.

## 2.2 Selection of the Papers

The papers were selected based on multiple factors, such as their relevance to this study, publication details, recency (published after 2018), uniqueness of methodological choices, and depth of findings. The fulfilment of these selection criteria for each selected paper is provided in the table below.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Title and Author** | **Relevance to Project** | **Machine Learning Focus** | **Customer Spending Focus** | **Food Products Context** | **Prediction/Forecasting** |
| “Applicability of machine learning techniques in food intake assessment: A systematic review” (Chaves *et al.*, 2021) | Explores machine learning techniques applicable to food assessment and relevant to predicting customer spending. | Strong focus on various machine learning methods using tools like WEKA and R. | Indirect:  Focus on food intake rather than spending. | Direct:  Applicable to food products. | Partial:  Focus on intake assessment rather than sales prediction using Supervised Learning methods. |
| “Predicting future consumer purchases in grocery retailing with the condensed Poisson lognormal model” (Trinh and Wright, 2022) | Highly relevant as it directly predicts consumer purchases in grocery retailing, similar to food product spending using ML techniques like Condensed Negative Binomial Distribution (CNBD) and Condensed Poisson lognormal model (CPLN) | Uses a Poisson lognormal model, which can complement ML approaches. | Direct focus on predicting consumer spending. | Direct:  Focus on grocery purchases, similar to food products. | Strong:  Predicts future purchases, aligning with spending. |
| “Customer Value Types Predicting Consumer Behavior at Dutch Grocery Retailers”(Janssens *et al.*, 2020) | Relevant for understanding how customer value types impact spending behaviour in grocery retailing. | Uses statistical and behavioural approaches, not primarily ML-based. | Direct focus on consumer behaviour and spending. | Direct: Focus on grocery retailing. | Strong: Predicts consumer behaviour and spending in retail. |
| “From intention to action: Predicting purchase behaviour with consumers’ product expectations and perceptions” (Kytö, Virtanen and Mustonen, 2019) | Relevant for understanding the link between consumer perceptions and purchase behaviour. | Focus on consumer behaviour prediction, applicable to ML frameworks. | Direct focus on predicting consumer behaviour. | Indirect: More focused on general purchase behaviour, less on food. | Strong: Predicts purchasing behaviour based on perceptions. |

***Table 1: Selection of the papers***

## 2.3 Critical analysis of key papers

**Paper 1:** “Applicability of machine learning techniques in food intake assessment: A systematic review” (Chaves *et al.*, 2021)

The research by Chaves *et al.* (2021) focused on the evaluation of applicability of the ML techniques in the estimation of food intake of customers. This work has utilised 36 secondary research on ML modelling for prediction of food consumption using Systematic Literature Review and Meta analyses. The obtained results of this study revealed that implication of Decision tree (DT), Support Vector Machine (SVM), Naive Bayes Regressor, k-Nearest Neighbour (KNN) Regressor and Artificial Neural Networks (ANN) are highly reliable in the prediction of customer food intake. Moreover, the obtained predictive accuracy of the DT and ANN model is considerably higher than other models, due to their capability of capturing non-linear relationships between food intake and customer demographics and spending habits. Moreover, the findings by Chaves *et al.* (2021) emphasised that the capability of the DT model to express possible results of a series of choices related to attributes (customer demographics and historical spending habits) through pre-defined criteria makes it a preferable choice for predicting food intake of customers.

The paper can be considered as partially linked with the motive of this project as it fails to provide detailed insights about customer spending on food products; instead, the paper has emphasised aspects of food intake preferences of customers. The good part of the paper includes detailed discussion on ML algorithms and computational tools for predicting food consumption. On the other hand, limitations include non-existence of predictive power of the ML for predicting food intake of customers and their possible applications in predicting food preferences of customers.

**Paper 2:** “Predicting future consumer purchases in grocery retailing with the condensed Poisson lognormal model” (Trinh and Wright, 2022)

The aim of this study was to predict future consumer purchases in grocery retailing based on attributes like previous purchase behaviour and customer demographics. The study by Trinh and Wright (2022) utilised two grocery retailing datasets from the UK and focused on the development of new models (which is a mixture of Erlang-2, Poisson distributions or a condensed Poisson lognormal model (CPLN)) for predicting future consumer purchase amount in grocery retailing. The implication of the CPLN and CNBD models is that they have reduced the errors in estimation of customer spending on grocery retailing by approximately 50% (7%) and 67% (8%) respectively, compared to benchmark models such as the Decision Tree (14%) or Linear regression (16%) (Trinh and Wright, 2022). The strengths of the study included a detailed discussion of the methodological framework followed and implication of the new approach for predicting customer spending on grocery retailing. Additionally, theoretical and practical implications for retailers have been discussed in this paper, which makes this study suitable for evaluating implications of the models within the field of retail management. On the other hand, the limitations included a lack of focus on traditional models such as DT, KNN, and SVM, which made the comparative analysis inadequate for this study.

**Paper 3:** “Customer Value Types Predicting Consumer Behavior at Dutch Grocery Retailers”(Janssens *et al.*, 2020)

The purpose of this paper was based on the prediction of customer value types (which are based on three different aspects such as Satisfaction, Repurchase Intention, and Word-of-Mouth) within the grocery retail sector in the Netherlands. The study by Janssens *et al.* (2020) used survey questionnaires as the research instrument and developed ‘Partial least squares structural equation modelling (PLS-SEM)’ for predicting factors affecting the value types of customers using SPSS Statistical Software. The results obtained from this study revealed that factors like low prices, interesting bargains, and easy access to the products to the customers have a substantial positive influence on high customer spending. This study is relevant to this research as it provided a theoretical base for evaluating factors affecting customer spending on food products. The limitations include a lack of focus on customer spending on food products rather, this study has focused on customer value types. Moreover, implication of ML algorithms for the prediction of customer spending is missing, which can show limitations of the approach in predicting customer behaviour based on large volumes of customer spending data. On the contrary, the strengths of this study include a detailed discussion of different aspects of customer value types within grocery retail formats (on-discounter, soft discounter, as well as hard discounter).

**Paper 4:** “From intention to action: Predicting purchase behaviour with consumers’ product expectations and perceptions” (Kytö, Virtanen and Mustonen, 2019)

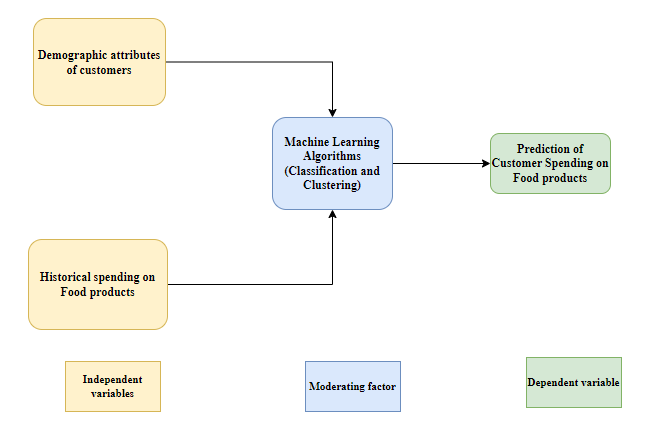
The study by Kytö, Virtanen and Mustonen (2019) explored the relationship between purchase intentions and actual purchase behaviour for two types of spoonable dairy snack products, which are natural yoghurt and flavoured protein quark. The aim of this paper was on the evaluation of two key phases of the buying process, which include expectations based on brand and package picture and perception after tasting the product at home. Within this study, data was collected through surveys involving two groups of consumers: users of natural yoghurt (n = 105) and users of flavoured protein quark (n = 107) from Finland. Based on the collected data from a survey of customers from the Food sector in Finland, predictive modelling (Logistic Regression) was performed to assess how well purchase intention ratings (both expectation and perception) could forecast actual purchase behaviour. Logistic regression models were developed to classify buyers and non-buyers for both product types, and an accuracy of 67.3% for yoghurt and 72.3% for quark was obtained from the model. The findings of this can be relatable to this study as it has explored consumer purchase behaviour in the context of food products, providing insights into how intention and perception influence actual purchases.

Main strengths of this study include the incorporation of multiple surveys and follow-ups with surveyed customers, ensuring a longitudinal view of consumer behaviour, which strengthens the predictive accuracy of the models. On the other hand, the limitations of this study include the lack of focus on diverse food products, which can limit the generalisability of the model in predicting food consumption.

## 2.4 Comparison of the papers

The paper by Chaves et al. (2021) has provided a detailed view of the implication of Ml algorithms on the prediction of food intake of customers based on the SLR approach. On the other hand, the study by Trinh and Wright (2022); Kytö, Virtanen and Mustonen (2019) have developed different state-of-the-art ML algorithms such as DT, KNN, SVM along with advanced techniques like condensed Poisson lognormal model (CPLN)) to predict customer behaviour. The obtained results revealed that integration of a large volume of customer demographic data and historical consumer spending allows businesses to predict customer spending on food products, which enables them to streamline their inventory level. Based on the comparative analysis, it can be stated that the study by Trinh and Wright (2022) has provided better insights related to the spending of customers on retail food products.

## 2.5 Conceptual Framework

******

***Figure 2: Conceptual Framework***

## 2.6 Literature Gaps

Findings of past studies have primarily focused on the evaluation of customer purchase intention and value types within the retail food sector. The studies by Trinh and Wright (2022); Kytö, Virtanen and Mustonen (2019) have shown a lack of focus on prediction of customer spending on food products. Moreover, a discussion of previous spending on food products and their impact on future spending has not been adequately provided within these studies. In fact, in previous studies, segmentation of customers based on purchasing power and demographics has not been adequately explored, which reflects a potential gap in past literature. In this context, this study has specifically focused on the implication of ML (Classification and clustering) techniques for evaluating customer spending on food products.

## 2.7 Summary

This chapter has helped in understanding the key research which has been conducted in understanding customer spending on food products with the help of machine learning. With the help of this chapter, it can be found that machine learning which includes both classification and clustering has not been done before which has helped in motivating in doing this research.

# References

Cai, J. (2023) ‘Research on the Influencing Factors of Consumer Buying Behavior’, *Highlights in Science Engineering and Technology*, 61, pp. 119–127. doi: https://doi.org/10.54097/hset.v61i.10281.

Chaves, L. O. *et al.* (2021) ‘Applicability of machine learning techniques in food intake assessment: A systematic review’, *Critical Reviews in Food Science and Nutrition*, 63(7), pp. 1–18. doi: https://doi.org/10.1080/10408398.2021.1956425.

Christy, A. J. *et al.* (2018) ‘RFM Ranking – an Effective Approach to Customer Segmentation’, *Journal of King Saud University - Computer and Information Sciences*, 33(10). doi: https://doi.org/10.1016/j.jksuci.2018.09.004.

Cognitive Market Research (2023) *Consumer Food and Beverages Market Research Reports*, *Cognitive Market Research*. www.cognitivemarketresearch.com. Available at: https://www.cognitivemarketresearch.com/list/food-%26-beverages/consumer-food-and-beverages (Accessed: 23 October 2024).

Daroch, B., Nagrath, G. and Gupta, A. (2021) ‘A study on factors limiting online shopping behaviour of consumers’, *Rajagiri Management Journal*, 15(1), pp. 39–52. doi: https://doi.org/10.1108/ramj-07-2020-0038.

Heldt, R., Silveira, C. S. and Luce, F. B. (2019) ‘Predicting customer value per product: From RFM to RFM/P’, *Journal of Business Research*, 127, pp. 444–453. doi: https://doi.org/10.1016/j.jbusres.2019.05.001.

Janssens, K. *et al.* (2020) ‘Customer Value Types Predicting Consumer Behavior at Dutch Grocery Retailers’, *Behavioral Sciences*, 10(8), p. 127. doi: https://doi.org/10.3390/bs10080127.

Kytö, E., Virtanen, M. and Mustonen, S. (2019) ‘From intention to action: Predicting purchase behavior with consumers’ product expectations and perceptions, and their individual properties’, *Food Quality and Preference*, 75, pp. 1–9. doi: https://doi.org/10.1016/j.foodqual.2019.02.002.

Statista (2024) *Food - Worldwide | Statista Market Forecast*, *Statista*. www.statista.com. Available at: https://www.statista.com/outlook/cmo/food/worldwide (Accessed: 23 October 2024).

Trinh, G. and Wright, M. J. (2022) ‘Predicting future consumer purchases in grocery retailing with the condensed Poisson lognormal model’, *Journal of Retailing and Consumer Services*, 64, p. 102809. doi: https://doi.org/10.1016/j.jretconser.2021.102809.