

Chapter 43

Identifying Patterns in Fresh Produce Purchases: The Application of Machine Learning Techniques

Timofei Bogomolov

University of South Australia, Australia

Malgorzata W. Korolkiewicz

University of South Australia, Australia

Svetlana Bogomolova

 <https://orcid.org/0000-0003-4449-6514>

Business School, Ehrenberg-Bass Institute, University of South Australia, Australia

ABSTRACT

In this chapter, machine learning techniques are applied to examine consumer food choices, specifically purchasing patterns in relation to fresh fruit and vegetables. This product category contributes some of the highest profit margins for supermarkets, making understanding consumer choices in that category important not just for health but also economic reasons. Several unsupervised and supervised machine learning techniques, including hierarchical clustering, latent class analysis, linear regression, artificial neural networks, and deep learning neural networks, are illustrated using Nielsen Consumer Panel Dataset, a large and high-quality source of information on consumer purchases in the United States. The main finding from the clustering analysis is that households who buy less fresh produce are those with children – an important insight with significant public health implications. The main outcome from predictive modelling of spending on fresh fruit and vegetables is that contrary to expectations, neural networks failed to outperform a linear regression model.

DOI: 10.4018/978-1-6684-6291-1.ch043

INTRODUCTION

Recent advances in technology have led to more data being available than ever before, from sources such as climate sensors, transaction records, scanners, cellphone GPS signals, social media posts, digital images, and videos, just to name a few. This phenomenon is referred to as Big Data, allowing researchers, governments, and organizations to know much more about their operations, thus leading to decisions that are increasingly based on data and analysis, rather than experience and intuition (McAfee & Brynjolfsson, 2012).

Big Data is typically defined in terms of its variety, velocity, and volume. Variety refers to expanding the concept of data to include unstructured sources such as text, audio, video, or click streams. Velocity is the speed at which data arrives and how frequently it changes. Volume is the size of the data, which for Big Data typically means large, given how easily terabytes to zettabytes of information are amassed in today's marketplace.

When it comes to consumer behavior and decisions, consumer data makes it possible to track individual purchases, to capture the exact time at which they occur, and to track purchase histories of individual customers. This data can be linked to demographics, advertising exposure, or credit history. Hence, researchers now have access to much more consumer data with greater coverage and scope, but also much less structure or much more complex structure than ever before. Traditional econometric modelling generally assumes that observations are independent, grouped (panel data), or linked by time. However, the Big Data we now have available may have more complex structure, and the goal of modern econometric modelling could be to uncover exactly what the key features of this dependence structure are (Einav & Levin, 2014). Developing methods that are well suited to that purpose is a challenge for researchers.

This chapter examines consumer food choices, in particular, purchasing patterns in relation to fresh fruit and vegetables. Consumption of fresh fruit and vegetables makes an important contribution to society in multiple ways. Increased consumption of fruit and vegetables can have a significant positive effect on population health (Mytton, Nnoahim, Eyles, Scarborough, & Mhurchu, 2014; World Health Organization 2015). Strong sales of fresh produce support primary production, contributing to rural and regional economies and farmers' livelihoods (Bianchi & Mortimer, 2015; Racine, Mumford, Ladditka, & Lowe, 2013). Fruit and vegetable categories in supermarkets contribute some of the highest profit margins, compared to other product categories (e.g., packaged food), making these categories very important for supply-chain members. Therefore, better understanding and prediction of patterns of consumer purchases of fresh fruit and vegetables could have a substantial positive effect on a range of health, economic, commercial, and social outcomes.

Traditionally, consumer research into fresh fruit and vegetables has relied on consumer surveys, where consumers report their attitudes and intentions to buy fresh produce and barriers to doing so (Brown, Dury, & Holdsworth, 2009; Cox et al., 1996; Péneau, Hoehn, Roth, Escher, & Nuessli, 2006; Finzer, Ajay, & Ali, 2013; Erinosh, Moser, Oh, Nebeling, & Yaroch, 2012). The results were inherently biased by the indirect link between what consumers say in surveys and their actual behavior. When fresh produce purchases were examined, they were often based on self-reports, which typically are influenced by social desirability bias (Norwood & Lusk, 2011) and memory failures, resulting in over- or under-reporting of purchases (Ludwichowska, Romaniuk, & Nenycz-Thiel, 2017). Overcoming these limitations, this chapter draws on a more reliable Consumer Panel Dataset, which is one of the Nielsen datasets made available to marketing researchers around the world at the Kilts Center for Marketing, the University of Chicago Booth School of Business. Since participating households routinely scan all their purchases,

29 more pages are available in the full version of this document, which may be purchased using the "Add to Cart" button on the publisher's webpage:

www.igi-global.com/chapter/identifying-patterns-in-fresh-produce-purchases/307485

Related Content

Automobile Predictive Maintenance Using Deep Learning

Sanjit Kumar Dash, Satyam Raj, Rahul Agarwal and Jibitesh Mishra (2021). *International Journal of Artificial Intelligence and Machine Learning* (pp. 1-12).

www.irma-international.org/article/automobile-predictive-maintenance-using-deep-learning/279274

Convolution Neural Network Architectures for Motor Imagery EEG Signal Classification

Nagabushanam Perattur, S. Thomas George, D. Raveena Judie Dolly and Radha Subramanyam (2021). *International Journal of Artificial Intelligence and Machine Learning* (pp. 15-22).

www.irma-international.org/article/convolution-neural-network-architectures-for-motor-imagery-eeeg-signal-classification/266493

Multi-Objective Materialized View Selection Using Improved Strength Pareto Evolutionary Algorithm

Jay Prakash and T. V. Vijay Kumar (2019). *International Journal of Artificial Intelligence and Machine Learning* (pp. 1-21).

www.irma-international.org/article/multi-objective-materialized-view-selection-using-improved-strength-pareto-evolutionary-algorithm/238125

An Integrated Process for Verifying Deep Learning Classifiers Using Dataset Dissimilarity Measures

Darryl Hond, Hamid Asgari, Daniel Jeffery and Mike Newman (2021). *International Journal of Artificial Intelligence and Machine Learning* (pp. 1-21).

www.irma-international.org/article/an-integrated-process-for-verifying-deep-learning-classifiers-using-dataset-dissimilarity-measures/289536

COVID-19-Related Predictions Using NER on News Headlines

Bidyut B. Hazarika, Urvi Trivedi, Harshita Dahiya, Nishtha Nandwani and Aakriti Gupta (2021). *Handbook of Research on Machine Learning Techniques for Pattern Recognition and Information Security* (pp. 187-208).

www.irma-international.org/chapter/covid-19-related-predictions-using-ner-on-news-headlines/279912