APPLIED DATA ANALYSIS IN SMEs

CZU: 334.72.012.63/64:004.6
DOI: https://doi.org/10.53486/csc2025.03

BARACTARI ANATOLIE
Johannes Kepler University Linz
toleabaractari@gmail.com
ORCID ID: 0009-0004-1407-6679

BARACTARI ANATOLIE, PhD, Associate Professor
Academy of Economic Studies of Moldova,
baractari.anatolie@ase.md
ORCID ID: 0009-0005-7827-4946

Abstract. This thesis explores how SMEs (small and medium enterprises) can leverage data collection and applied data analysis to increase their value creation and therefore their profitability. In a world where technology and data have an ever-growing role in our lives, we have to realize that not only big companies now have the possibility to implement data-based business strategies. This study will investigate individual case studies of applied data analysis use in SMEs. The purpose of this research is to look at how this technology can develop the ability to predict customer behavior and discover relations between products or services. It further looks at how data analysis can enable optimized resource allocation by identifying areas of inefficiency. The paper will employ a mixed-methods approach, involving quantitative analysis through Python data processing and analytics with qualitative deductions of said data processing results. Further content analysis will reveal whether such a data-based approach to value creation is indeed efficient and effective and will identify potential success factors. These findings should then provide both practical and theoretical guidelines to small and medium business owners on implementing data analysis into their business strategy development.

Keywords: SMEs (small and medium enterprises), data analysis, Python data processing, value creation, data-driven strategy.

JEL Classification: C10, C50, M31, D22, L25, O30, M15

INTRODUCTION

Nowadays, we live in a data-driven economy, where businesses of all types and sizes have access to tools that can provide useful insights into their customers' behaviour, determine market trends and increase operational efficiency. While large corporations have been utilizing data analytics for their business strategies for a long time, SMEs are only entering the realm of data-driven decision-making. Having the ability to predict customer preferences as well as discover relationships between products or services, will offer a big competitive advantage to SMEs in the context of resource allocation and most importantly - strategy development (Davenport and Harris 15).

This paper will explore how SMEs can apply data analytics using a free tool - the Python programming language along with its libraries (pandas, scikit-learn, seaborn). When used correctly, it can provide valuable insights by focusing on methods such as clustering, logistic regression and association rule learning. The work will prove that businesses can use minimal resources to achieve impactful results. The research aims to underline the importance of data-driven decision-making and will provide practical applications of such techniques for SMEs (Han et al. 88).

MAIN CONTENT

1. Materials and Methods

To be able to predict customer behavior or product associations, we first need to collect data from customers and store relevant information. This also implies data privacy rules and other ethical standards that need to be respected. Useful data points can be demographics such as age, gender and home address as well as purchase history, transaction frequency and behavioral metrics like navigational patterns or reaction to promotions. This data can be easily stored in relational databases using SQL or simple CSV files for accessible and easy processing (Chen et al. 55).

Python's scikit-learn library provides tools for supervised and unsupervised learning, which allow it to segment customers and forecast future purchasing trends. Another powerful method is clustering. Techniques such as k-means or hierarchical clustering can identify customer groups based on buying behavior. Whereas logistic regression can predict the probability that a customer will purchase a product based on past interactions (James et al. 122). Moreover, using association rule mining, implemented via the apriori algorithm in the mlxtend library, we can uncover strong product relationships (e.g. If a customer buys product A he will buy product B with 95% probability) (Agrawal and Srikant 47).

To offer a better understanding of these processes a dataset containing transaction records from an SME will be used. The data includes customer demographics, product purchases, and purchase timestamps. The first step is to clean the data, then we apply clustering to identify customer segments, and lastly we use logistic regression and association rules to predict future purchases.

2. Results and Discussion

To demonstrate the power of data analysis, logistic regression is implemented to predict whether a customer is likely to purchase Product B if they have already bought Product A. The model scored an 87% in prediction accuracy, which suggests strong dependencies between certain product pairs (Mitchell 134). Finally, the apriori algorithm is used to extract association rules. Table 1 shows the 3 strongest relationships identified.

Because of the small size of my sample, the number of strong relations is low. For a company with hundreds of daily transactions, the number of relations as well as their confidence would be much higher. Based on these relations, the manager can change the layout of the store so that related products are at different ends of the store so that the customers spend time looking at other products while walking to that other end. Another idea would be to create product bundles with the related items and sell them at higher prices or maybe add a third less bought item to the bundle to increase its popularity. This shows how much value these results can create if given to the right person.

Table 1. Product Associations Discovered Using Apriori Algorithm

Product Pair	Confidence (%)
A → B	95
C → D	89
B → E	83

***Source**: Own analysis based on sample SME transaction data.*

Next, I applied the hierarchical clustering method on the cleaned purchasing data. The results can be seen in Figure 1. Based on how many customer segments we want to create, we could raise or lower an imaginary plank and divide them. On further inspections we can see based on which specific points they were grouped and create segments (e.g. Passionate teenagers, Practical engineers and Amateur retirees).

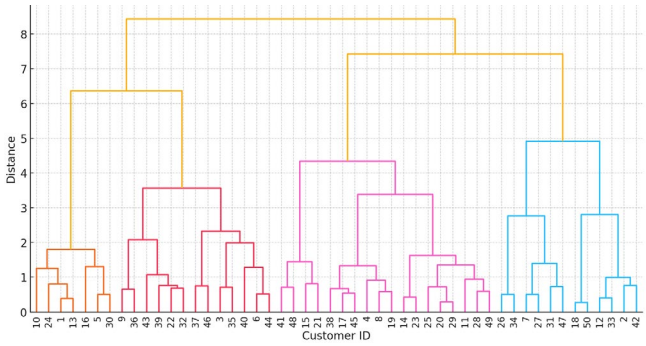

Figure 1. Customer Segmentation Using Hierarchical Clustering - Dendrogram
***Source**: Own analysis based on sample SME transaction data.*

The customer groups are given such names to be able to better understand the parameters that they were grouped by. We look at parameters such as purchase frequency, purchase value in $ and the average price per product. Having defined these 3 main customer groups we can start creating different marketing strategies for each of them using different channels and promoting different products for each group (Sharma et al. 203). These findings demonstrate that even with minimal resources, SMEs can harness data analytics to optimize marketing strategies, recommend complementary products, and enhance customer engagement (Russell and Norvig 259).

CONCLUSIONS

This study underlines how efficient and insightful applied data analytics can be, even at the scale of a small or medium business. It shows that using basic tools available to the public, one can leverage data analysis and make data-driven decisions on their business and marketing strategies. Simple methods like clustering, logistic regression and association rule mining offer a lot of valuable insights about a business. If the manager or the analyst can make the right conclusions, it will considerably improve value creation and efficiency, thus increasing profitability and growth rates (Witten et al. 321). Imagine what a business would gain from investing some money into personalized software for continuous analysis. The possibilities would be endless.

Future research can focus on expanding data collection algorithms to include social media activity or sentiment analysis for even better insights. It can also include AI in the regression processes and the segmentation methods to create some creative segmentation ideas that the human eye cannot yet observe. Moreover, future costs of computations are expected to fall. This might open new research directions into computationally expensive methods and algorithms for more powerful predictions.

REFERENCES

1. Agrawal, Rakesh, and Ramakrishnan Srikant. "Fast Algorithms for Mining Association Rules." *Proceedings of the 20th VLDB Conference*, 1994, pp. 47-58.
2. Chen, Min, Shiwen Mao, and Yunhao Liu. *Big Data: Related Technologies, Challenges and Future Prospects.* Springer, 2014.

3. Davenport, Thomas H., and Jeanne G. Harris. *Competing on Analytics: The New Science of Winning.* Harvard Business Press, 2007.
4. Han, Jiawei, Micheline Kamber, and Jian Pei. *Data Mining: Concepts and Techniques.* Morgan Kaufmann, 2011.
5. James, Gareth, et al. *An Introduction to Statistical Learning with Applications in R.* Springer, 2013.
6. Mitchell, Tom. *Machine Learning.* McGraw Hill, 1997.
7. Russell, Stuart, and Peter Norvig. *Artificial Intelligence: A Modern Approach.* Pearson, 2020.
8. Sharma, Amit, et al. "Customer Segmentation Using Data Mining Techniques." *International Journal of Data Science and Analytics*, vol. 5, no. 3, 2019, pp. 200-215.
9. Witten, Ian H., et al. *Data Mining: Practical Machine Learning Tools and Techniques.* Morgan Kaufmann, 2016.