

Metoda klasterizacije u sektoru modne industrije u cilju podizanja kvaliteta poslovanja za donosioce odluka

dr Stefana Janićijević¹
dr Svetlana Milutinović²
Nenad Avlijaš³

Apstrakt:

Rad prikazuje uticaj pametnih tehnologija na donošenje odluka u kompanijama širom sveta. Fokusiran je na pružanju podrške modnoj industriji i marketinškim odlukama baziranim na pristupu podataka kroz Data Mining (DM). Današnji korisnici imaju tako raznovrsne ukuse i želje da ih nije moguće grupisati u velike i homogene populacije kako bi razvili marketinške strategije. Zapravo, svaki korisnik želi da bude servisiran prema svojim individualnim i jedinstvenim potrebama. Prema tome, prelazak sa masovnog marketinga na tržište odnosa jedan-na-jedan zahteva donosioce odluka da daju posebne strategije za svakog pojedinačnog korisnika na osnovu njegovog profila. Tehnologija budućnosti pa i sadašnjosti je veštačka inteligencija koja se primenjuje uspešno u svim sferama poslovnih procesa. Jedna od najčešće korišćenih metoda veštačke inteligencije je metoda pametne klasifikacije, odnosno klasterovanja, bez uticaja pristrasnih faktora čoveka. Za istraživanje je izolovana baza iz maloprodajnog sektora gde se posmatrala istorija kupovina lojalnih korisnika i gde se primenio algoritam klasterovanja iz oblasti inteligentnih sistema ili veštačke inteligencije.

Ključne reči: donošenje odluka, marketing, menadžment, veštačka inteligencija, modna industrija

¹ ASW Inženjering, Takovska 45A, stefana@mi.sanu.ac.rs
² Telekom Srbija, Dubrovačka 35, svetlanamilu@gmail.com
³ ASW Inženjering, Takovska 45A, nenad.avlijas@asw.eu

Method of clustering in fashion industry sector with the aim of raising the quality of business for the decision makers

Abstract:

The paper presents the influence of smart technology on decision making in companies around the world. It's focused on the problem of supporting fashion industry and marketing decisions based on Data Mining (DM) approaches. Today's customers have such a varied tastes and preferences that it is not possible to group them into large and homogeneous populations to develop marketing strategies. In fact, each customer wants to be served according to his individual and unique needs. Thus, the move from mass marketing to one-to-one relationship marketing requires decision-makers to come up with specific strategies for each individual customer based on his profile. The technology of the future and the present is artificial intelligence that is applied successfully in all spheres of business processes. One of the most commonly used methods of artificial intelligence is the method of smart classification, i.e. clustering, without the influence of biased factors of man. The research is an isolated base from the retail sector where the history of buying loyal customers was observed and where clustering algorithm applied in the field of intelligent systems or artificial intelligence.

Key words: decision making, marketing, management, artificial intelligence, fashion industry

Introduction

Modern society, whose main characteristic is data, has been in the Internet revolution for several decades now. Companies such as Google, Facebook, Amazon, Microsoft, who first realized this, took the world's first development and profits and became the state-of-the-art company that everyone sees. The methodology of data warehousing and manipulation is extremely important as well as the methodology of

data interpretation. This analytics is not just a presentation in spreadsheets, but it is a complete communication technology with supercomputers and supercomputers, in combination with sensors, cloud, robotics, etc. Data research dates back to the beginning of the twentieth century with the first development of computing, then caused by wars of research turning to patches such as cryptography and logic.

The next point in history that is significant for the development of business data is the 70s and 80s of the last century. Nevertheless, it is worth mentioning recent sources such as McKinsey's work from the 2012 World Economic Forum in Davos, Switzerland, where he presented a survey titled "Big Data, Big Impact". After the aforementioned work, Barack Obama's Cabinet proclaimed the Big Data Initiative for which \$ 200M was allocated and where institutions such as the Academy, the Government, NGOs and private corporations were invited to work on exploring key points on the big data.

Today, the situation is such that data takes an essential role in people's lives. The term Big Data has gone down in history and now the brief data of all forms is called Data.

There are three relevant terms in science and technology that are direct participants in modern marketing and decision making. It is almost unthinkable to create a successful business in the future without the following three components: [1]

- Data mining (DM) - the discipline of As the purpose of modern business is growth in every direction, the demands placed on new technologies are becoming more and more challenging discovering pattern in large data sets; baseline for analysis of all types in structured and unstructured data; It is based on database systems and statistics
- Machine Learning (ML) - discipline of using algorithms that create predictions from historical data, whereby, depending on the amount of historical data, the computer trains to be independent
- Artificial Intelligence (AI) - a broad discipline that includes data mining and machine learning; uses processes from these disciplines to describe the characteristics of human intelligence and at some point with sufficiently technologically developed societies the question of the role of traditional marketing in business is posed.

Fashion industry and decision making

Fashion is a major global industry. The global apparel market is valued at between US\$2.4 trillion^{2/} and US\$3 trillion and accounts for 2 percent of the world's Gross Domestic Product (GDP). 57.8 million People are employed in clothing and textiles worldwide – 24.8 million of those in apparel manufacture. [2]

Digital technologies reshape markets and value chains for fashion content and information, leading to an opportunity for innovative businesses to create value added services, applications and products. ICT forms the enabling element that brings to market these services, applications and products across all sectors through production, distribution and e-commerce. The life cycle of fashion products becomes shorter in recent years due to the fierce market competition environment. Short life cycles, high volatility, low predictability, and high impulse purchasing is being appointed has characteristics of fashion industry. The data that companies collect about their customers is one of its greatest assets. However, companies increasingly tend to accumulate huge amounts of customer data in large databases and within this vast amount of data are all sorts of valuable information that could make a significant difference to the way in which any company run their business, and interact with their current and prospective customers and gaining competitive edge on their competitors. [3] Global markets demand innovation.

DM is a very powerful tool that should be used for increasing customer satisfaction providing best, safe and useful products at reasonable and economical prices as well for making the business more competitive and profitable. A clustering algorithm assigns data points to different groups, some that are similar and others that are dissimilar. The use of clustering involves placing data into related groups typically without advance knowledge of group definitions. Companies can utilize DM techniques to extract the unknown and potentially useful information about customer characteristics and their purchase patterns DM tools can, then, predict future trends and behaviours, allowing businesses to make knowledge driven decisions that will affect the company, both short term and long term. The identification of such patterns in data is the first step to gaining useful marketing insights and making critical marketing decisions. [4]

The concept of applied statistics in the sales sector

The first question that concerns business and the data is - How to use Data Analytics to increase the shareholder value?

Consumer society based business relies on various media that serve as a channel for data integration, where both structured and unstructured data are rapidly growing day by day. Companies have billions of numerical records, demographics, records from social networks, textual records, video and image records, etc. Such a conglomerate base is a challenge for successful business that balances between strategy, marketing, structures, and algorithms. The business that first sees alarms, anomalies, rules, and results in data has the lead in an everlasting competition between the competitions.

Cluster analysis or clustering is the task of grouping customers in such a way that customers in the same cluster group are more similar to customers in other groups. Customers in the same clusters share features that lead to a similar product group or similar service group.

Algorithms as mathematical concepts are aimed at theoretical grounding, however, their expansion comes only after application in the economy and industry. The science of data that has evolved over the years gets its shape only when it is applied to concrete datasets and when results that yield success on it are seen.

Many algorithms are developed for the benefit of data science; it is worth mentioning only the most basic groups: Clustering, Logistic regression, Linear models, Support vector machines, decision trees, neural networks, etc.

This paper presents clustering. Clustering is one of the methods of unexpected learning that allows clustering of instances into groups, where the number and size of groups is not known in advance. The essence of clustering is to create as homogeneous groups as possible as heterogeneous among themselves. Clustering goals are a better understanding of consumer behaviour, sales campaign planning, customer retention, optimization of marketing costs, quality engagement of customers through appropriate channels and impact on consumer behaviour. The clustering information can be used to “tag” customers in the overall database. Customer clustering uses purchase transaction data to track buying behaviour and then create new business

initiatives based upon findings, like sales campaigns and customer retention. Cluster analysis is one of the most important segmentation methods and it has long been the dominant and preferred method for market segmentation. [5] Therefore, clustering methods are commonly used in marketing for the identification and definition of market segments that become a focus of a company's marketing strategy. [6] Traditionally, marketers must first identify customer cluster using a mathematical mode and then implement an efficient campaign plan to target profitable customers. [7]

Applications of Clustering

Clustering has a large number of applications spread across various domains. Some of the most popular applications of clustering are:

- Recommendation engines
- Market segmentation
- Social network analysis
- Search result grouping
- Medical imaging
- Image segmentation
- Anomaly detection

Customer clustering is the most important data mining methodologies used in marketing and customer relationship management (CRM). Customer clustering would use customer-purchase transaction data to track buying behaviour and create strategic business initiatives.

Being part of a cluster allows companies to operate more productively in sourcing inputs; accessing information, technology, and needed institutions; coordinating with related companies; and measuring and motivating improvement. [8]

Competition in today's economy is far more dynamic. Companies want to keep high-profit, high-value, and low-risk customers. This cluster typically represents the 10 to 20 percent of customers who create 50 to 80 percent of a company's profits. A company would not want to lose these customers, and the strategic initiative for the segment is obviously retention. A low-profit, high-value, and low-risk customer segment is also an attractive one, and the obvious goal here would be to increase profitability for this segment. Cross-selling (selling new products) and

up-selling (selling more of what customers currently buy) to this segment are the marketing initiatives of choice. [9]

K Mean algorithm

Clustering is widely used method which include huge number of algorithms. K Means algorithm was applied for the purpose of the research.

K Mean algorithm is a part of the Centroid method.

These are iterative clustering algorithms in which the notion of similarity is derived by the closeness of a data point to the centroid of the clusters. K-Means clustering algorithm is a popular algorithm that falls into this category. In these models, the number of clusters required at the end have to be mentioned beforehand, which makes it important to have prior knowledge of the dataset. These models run iteratively to find the local optima. [10]

Given a set of observations $(x_1, x_2, \dots, x_n)$ where each observation is a d -dimensional real vector, k -means clustering aims to partition the n observations into k ($\leq n$) sets $S = \{S_1, S_2, \dots, S_k\}$ so as to minimize the within-cluster sum of squares (WCSS).

Formally, the objective is to find:

$$\arg \min_S \sum_{i=1}^k \sum_{x \in S_i} \|x - m_i\|^2 = \arg \min_S \sum_{i=1}^k |S_i| \text{Var } S_i$$

where m_i is the mean of points in S_i . [9]

Methodology of research

Research was developed in ASW Engineering Company in Belgrade, Serbia, in Machine Learning department. The survey was carried out on the test set from the textile company's sales sector in the period from 2009 to 2017 years. On that occasion, the number of 1800 loyal customers (consumer participants) was isolated, out of which 500 significant (active) loyal customers were isolated. The number of purchases made in that period amounted to 57,100, while the total turnover per account of the account customers amounted to 1.600.000€.

- The total number of purchased items of loyal customers is 56.360

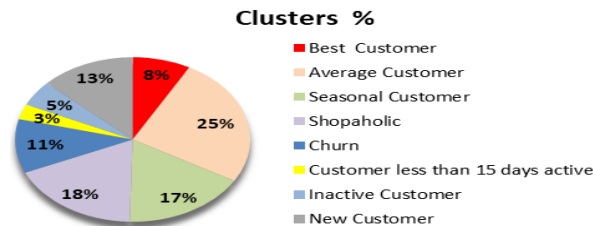
- Total number of purchases with a discount of these customers is 54.236

Based on the testing of multiple algorithms, it has been concluded that it is necessary to segment these customers into certain categories in order to make it easier for them to commune and create precise and successful campaigns.

The used characteristics or variables in the cluster model are divided into two groups - basic characteristics and special characteristics. The basic characteristics are the number of purchases so far (the number of days in which it was purchased), the average amount on the account, the length of the loyalty card, the frequency of purchases (average number of purchases per year). The specific characteristics are Vitality (the probability that the customer is active, calculated on the basis of a special churn model), Expected average amount in the account in 2018 (calculated by a special model), Expected number of purchases in 2018 (calculated by a special model), Expected revenue in 2018 (calculated by a special model).

After determining the variables that were involved in modelling, a kMean cluster was applied and 8 clusters were obtained, of which 6 are significant for analysis. Clusters are named on the basis of the most important characteristics: Average, Seasonal, New, Churn, Shopaholic, Best customers, and in addition are Inactive and customers who are less than 15 days active, shown on figure 1:

Figure 1: Percentage of clusters



Cluster 1 - Average

This cluster is characterized by buying cheap items and buying mostly women's things, such as hollyhocks and women's jeans. Shoes buy significantly less than average. The number of loyal customers in the cluster is 127, the number of men is 42, and the number of women is 85. This is the most numerous clusters, which accounts for 15.5% of

the total revenue of loyal customers. Customers of this cluster are more vital than average, and the length of possession of a loyalty card is also higher than the average of all loyal customers.

Cluster2 – Shopaholic

This cluster is characterized by buying it often; however, the amount on the accounts is less than the average. The cluster is predominantly women who have been loyal to the program for a long time. There is little risk of leaving these customers. Concerning revenue share, 35% of total revenue comes from this cluster. Expected revenue is significantly higher than the average in 2018.

Cluster3 – Seasonal

This cluster is characterized by buying an average of two times a year, in the summer and winter periods. Customers are at a low risk of becoming inactive (churn), most often buying jackets, caps and equipment for the sea, however the revenue predicted in 2018 is small.

Cluster4 – New

This cluster is characterized by buying up to half a year in a loyalty program. Mostly women make this cluster that buys most common earrings, bags, chains. This cluster is at a low risk of churn. This cluster predicts significantly higher revenue than average.

Cluster5 – Churn

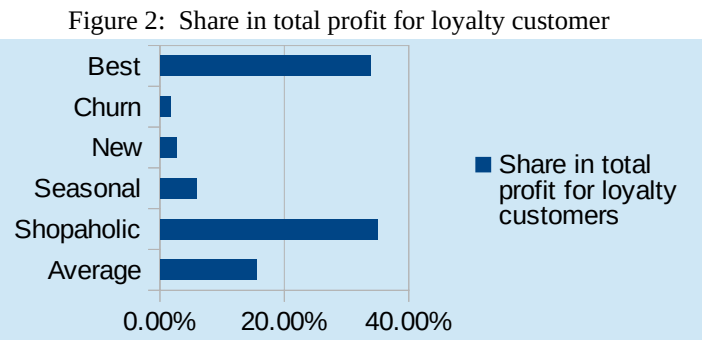
This cluster characterizes to buy shoes much more than other clusters. The likelihood that customers become inactive (churn) is as big as 50%. These are generally good buyers who have a slight downward trend in purchasing.

Saving a loyal customer is 10 times cheaper than gaining a new loyalty, it was necessary to pay extra attention to this cluster. It is divided into two groups, weaker outgoing customers and better outgoing customers. The properties that were used for additional clustering are vitality and expected revenue in 2018. The goal was to anticipate which customers are leaving and what the revenue from each customer is in 2018. The calculator has shown that if the departing loyal customers retain and thereby cross into the nearest clusters, they predict the income from 7.500 €.

Cluster6 – Best

This cluster is characterized by the purchase of extremely often and extremely expensive items. The account amounts for these customers are extremely high. They are naturally characteristic. They are expected to earn 3 times the average. About 34% of total income comes from the best buyers, shown on table 1 and figure 2. More often, they buy steaming items than individual goods.

Table1. Share in total profit for loyalty customers



Conclusion

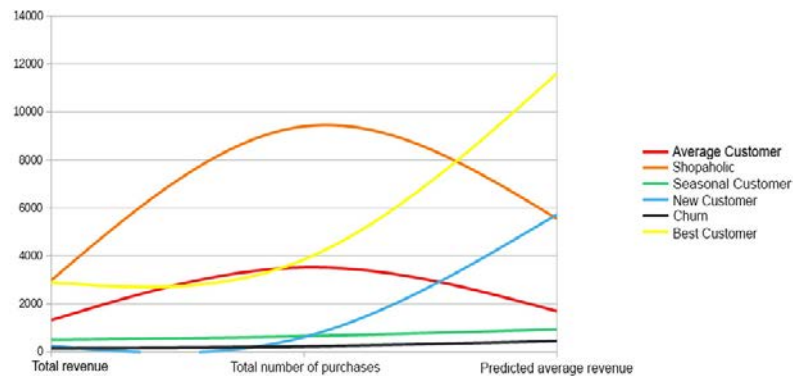
Clusters serve to make it easier to do upselling and cross selling. Similarities and differences are observed, as well as the hierarchy of clusters and activities are undertaken in the direction of migration of customers from the cluster to the cluster, in order to increase profitability and to influence the change in the behaviour of the consumer-participants.

Suggestion:

Average	→	Shopaholic (increase number of purchases)
Shopaholic	→	The Best (increase amounts on the account)
Seasonal	→	The Best (increase number of purchases and amount on the account)
New	→	Shopaholic (increase number of purchases)
Churn	→	Seasonal (increase number of purchases in the season)

Cluster hierarchy has made the following recommendations:

Figure 3: Cluster hierarchy



The figure 3. shows which cluster comes from and in which direction it works best in order to show the results first. The new buyers become Shopaholic when increasing the number of purchases, and on the other hand, seasonal buyers become an increase in the number of purchases by the average buyers. The average increase in the number of purchases is becoming a buyer for everyone. Shoppers, when buying expensive items, become the best buyers. Outgoing customers in a better group need to be kept in order to make them the Best Buyers.

In general, clustering as a technique of machine learning is extremely useful in high-positioned decision making, and this technique is basically modelling sales - missed, linked, personalized, targeted. Intelligently programmed groups are useful for BI analysis due to accurate customer comparison; Groups are useful and necessary when

designing group sales campaigns; groups introduce the company into a new level of business and build the level of personalization of customers; The groups give the possibility of alarm or easier detection when it becomes inactive (churn), who changes behaviour and how. Statistical modelling in company management processes and decisions in these companies is a sufficient and necessary condition for a sustainable and successful business.

Literatura

- [1] Hartley, M. (2005). *Using Data Mining to predict inventory levels*. Data Mining Book
- [2] Bakhshi, H., Mateo's-Garcia, J. (2016). *The Geography of Creativity in the UK Creative clusters, creative people and creative networks*. London: NESTA
- [3] Shaw, M.J., (2001). Knowledge management and data mining for marketing. *Decision support systems*,. 31(1): p. 127-137.
- [4] Krishna Murthy, I. (2010). *Data Mining- Statistics Applications: A Key to Managerial Decision Making*. Article/Report indiastat.com
- [5] Xie, W., Z. Huimin (2010). A talent markets analysis method based on data mining. in *Intelligent Computing and Intelligent Systems (ICIS), IEEE International Conference on Data Mining and Knowledge Management for Marketing*
- [6] Wedel, M., Kamakura, W. A. (2000). *Market Segmentation: Conceptual and Methodological Foundations*. Dordrecht: Kluwer. p.57
- [7] Lefait, G., Kechadi, T. (2010). *Customer Segmentation Architecture Based on Clustering Techniques*. Digital Society, ICDS'10, Fourth International Conference
- [8] Dubes, R. C. Jain, A. K. (1998). *Algorithms for Clustering Data*. Prentice Hall. p. 86
- [9] Mehrotra K., Mohan,C., Ranka, S. (1996). *Elements of Artificial Neural Networks*. MIT Press. p. 112
- [10] Fraley, A., Thearting, K. (1999). *Increasing customer value by integrating data mining and campaign management software*. *Data Management*, p.49–53.