THE USA JOURNALS
THE AMERICAN JOURNAL OF ENGINEERING AND TECHNOLOGY (ISSN – 2689-0984)
VOLUME 06 ISSUE10

PUBLISHED DATE: - 02-10-2024
DOI: - https://doi.org/10.37547/tajet/Volume06Issue10-04 PAGE NO.: - 23-32

RESEARCH ARTICLE Open Access

UNSUPERVISED MACHINE LEARNING AND VECTOR MODELS IN DESIGNING AND OPTIMIZATION OF TELECOM RETAIL CHANNELS

Andrei Zhuk
Manager, McKinsey & Company, New York
Wharton School of Business (University of Pennsylvania), Philadelphia, USA
MGIMO, Moscow, Russia

Abstract
This paper examines the use of unsupervised machine learning and vector models in the design and optimization of retail channels for telecommunications services. Unsupervised machine learning allows you to analyze and identify hidden patterns in large volumes of untagged data, which is especially important in a dynamically changing consumer market. Vector models, in turn, provide high accuracy of demand forecasting and inventory management, contributing to an increase in the efficiency of trading channels. The synergy of these technologies allows companies to improve customer experience, optimize operational processes and increase competitiveness in the market. The main focus of the work is on data processing methods, including correlation analysis, the use of the support vector machine (SVM) method and its adaptation to solve problems related to predicting customer behavior and optimizing logistics processes.

Keywords Uncontrolled machine learning, vector models, optimization of trade channels, telecommunication services, demand forecasting, support vector machine method, logistics, customer experience, competitiveness.

INTRODUCTION

In the context of rapid technological development, both retailers and telecom operators face challenges in managing physical sales networks, as maintaining a network of physical stores is one of the primary cost drivers. The efficiency of managing these stores is influenced by modern technological solutions. Traditional methods of data analysis and demand forecasting are becoming less effective due to the increasing volume and complexity of data. In this context, the use of unsupervised machine learning and vector models appears particularly promising.

Unsupervised machine learning enables the analysis of large volumes of unstructured data, uncovering hidden patterns and optimizing business processes without the need for pre-labeled data. This approach is especially relevant in the telecommunications sector, where consumer data and behavior play a key role in decision-making, making this topic highly relevant. The use of unsupervised machine learning and vector models significantly enhances operational efficiency, leading to cost reduction and improved customer experience. These technologies open new possibilities for analysis and forecasting, contributing to more accurate and faster decision-making.

The aim of this work is to explore the potential application of unsupervised machine learning and vector models for designing and optimizing retail channels for telecommunications services.

1. Application of Unsupervised Machine Learning in Consumer Data Analysis

In the process of studying and processing data within the framework of machine learning, especially when dealing with models that operate on information, it is crucial to use accurate and reliable data. Training models implies that algorithms must receive the necessary data to effectively perform tasks using artificial intelligence. While this approach is a popular method in machine learning, certain challenges arise with its use. One of the most significant challenges is the issue of data labeling, as finding accurately labeled data for feeding into the model is often difficult. Moreover, the cost of data can be high, and its use may not always yield the expected results. Currently, alternative methods are being developed that have not yet gained widespread use but may become important in the future.

One method worth noting is label-free learning. In this technique, the model receives raw data without predefined labels or patterns. The algorithm independently analyzes this data, generating new patterns and labels. One of the advantages of this approach is that there is no need to provide labeled data. The system autonomously identifies the rules necessary for analysis and generates its own patterns. The process of unsupervised learning involves several key stages, which are illustrated in Figure 1.

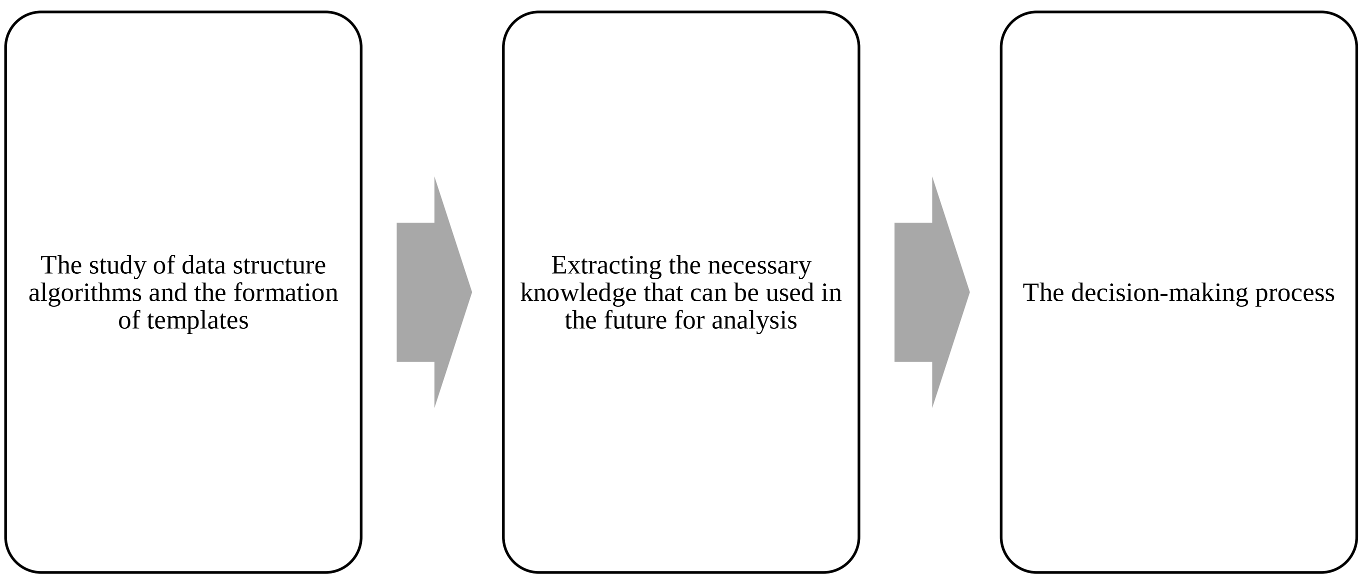

Fig.1. The stages of unsupervised learning [1].

Unsupervised learning algorithms can be classified into two main types based on the methods used for data processing, as shown in Figure 2.

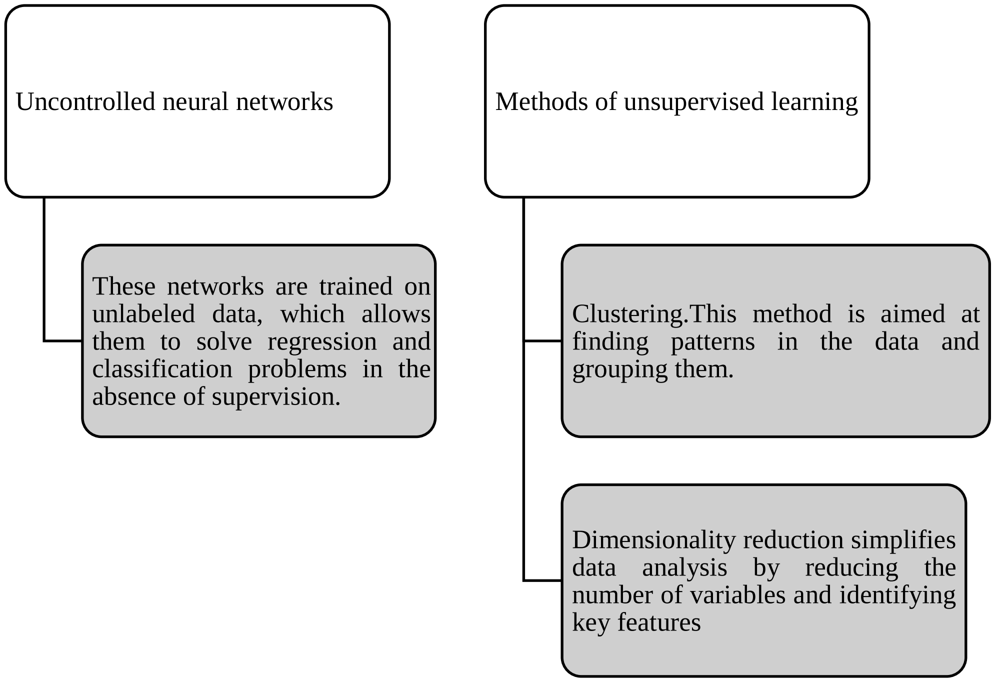

Fig.2. Classification of unsupervised learning algorithms [1].

Overall, unsupervised learning methods provide extensive opportunities for data analysis and processing, making them essential tools in the development of artificial intelligence [1]. The main distinction between supervised and unsupervised learning lies in the use of labeled data. Supervised learning relies on pre-known labeled input and output data, allowing the algorithm to learn by adjusting its predictions until the correct result is achieved. This requires human involvement and accurate data labeling. An example of this would be predicting commute times based on various factors, where the model is first trained on historical data [2].

There are other differences between these approaches. The goal of supervised learning is to predict outcomes for new data based on known patterns. In unsupervised learning, the focus is on uncovering hidden information from a large volume of new data.

Supervised learning is suitable for classification tasks, such as determining spam, sentiment analysis of texts, or weather forecasting. Unsupervised learning, in turn, is useful for tasks related to anomaly detection, improving recommendation systems, predicting customer behavior, and analyzing medical images [3].

2. Vector Models in Demand Forecasting and Distribution Channel Optimization

To obtain accurate data on future customer churn, a deep understanding of classification tasks, which are part of data mining, is required. The essence of this task is to create a mathematical model based on historical customer data to predict their behavior using the available information. In this forecast, customers are divided into two classes, and the term "label" is used to designate them, which can take two values:

- High probability of customer churn (churn).

- High probability of continuing service usage [4].

An important stage before training is data preparation, including cleaning outliers and eliminating duplicate features. Correlation analysis helped identify and remove redundant features, improving the model's quality. To test the model's accuracy, the data was split into training and test sets. As a result, it was found that the model's accuracy reached 92%, confirming the high effectiveness of using the support vector machine (SVM) method in predicting customer churn.

Vector data models are a powerful tool for analyzing and optimizing business processes, including logistics and inventory management in a retail network. These models represent objects and their characteristics as multidimensional vectors, enabling a more accurate analysis of the relationships between different parameters. In this context, vector model construction methods mainly include factor analysis, clustering, and neural networks.

Factor analysis is used to reduce data dimensionality and highlight the key factors that most influence stock dynamics and logistical processes. This allows for models that focus on the most significant characteristics, reducing excess information and simplifying the optimization process.

Neural networks, in turn, are used for demand forecasting based on historical data, a key element in inventory management. Recurrent neural networks (RNNs) and their modifications, such as LSTM, can model complex temporal dependencies, making them effective tools for predicting the demand for specific products in the retail network.

The practical application of vector models for optimizing logistics and inventory management manifests in several areas. First, they allow for more accurate forecasts of inventory needs, reducing the risk of overstocking or understocking. Second, these models help optimize delivery routes, minimizing time and logistics costs. Third, vector models can be used to automate inventory management processes, reducing labor costs and increasing decision-making accuracy.

Additionally, a program implementing an SVM classifier (support vector machine used for classification and regression analysis tasks) was tested for accuracy and performance speed. The conclusions drawn from the analysis indicate that the model can be successfully used to classify new customers, with a high degree of confidence in its predictions [5].

Thus, vector models represent an effective tool for improving logistics and inventory management in the retail network, providing more accurate forecasting, process optimization, and automation of routine operations.

3. Interaction of Unsupervised Machine Learning and Vector Models in Designing Sales Channels

One of the key methods of unsupervised machine learning applied in this field is clustering. Clustering allows grouping data based on similar features, identifying segments of consumers, product types, or geographic regions with shared characteristics. In this context, vector data models, which represent objects as multidimensional vectors, serve as the foundation for conducting cluster analysis. For example, customer profiles, represented as vectors, can be divided into clusters, enabling more precise adjustments to marketing strategies and sales offers for different audience segments [6].

Another important area of applying unsupervised machine learning in combination with vector models is cost reduction and logistics optimization. Dimensionality reduction algorithms, such as principal component analysis (PCA), can be used to simplify multidimensional vector models that represent various logistics process parameters. This simplification helps identify key factors influencing the efficiency of sales channels, such as delivery speed, inventory levels, or transportation routes. The patterns identified are then used to design more efficient logistics systems.

Neural networks that utilize unsupervised learning, such as autoencoders, also find application in the design of sales channels. These models are capable of identifying latent data representations that are not obvious using traditional analysis methods. Vector models trained with autoencoders can assist in automatically detecting anomalies in sales processes or predicting changes in consumer behavior, allowing sales channels to adapt to new market conditions and customer needs [7].

Thus, the interaction of unsupervised machine learning and vector models in the design of sales channels provides deeper data analysis and opens up opportunities for adaptive management of sales processes. This interaction not only reveals hidden patterns and structures in the data but also utilizes them to create more efficient, flexible, and resilient sales channels.

4. Practical Experience

As part of professional activities, an advanced analytical vector model was created to assess the potential of micro-markets based on both internal and external data, for both existing and new stores. An information dashboard was also developed to display the potential of individual stores and compare them with similar stores in the network. As a result, actual gross profit from sales in 245 new telecom retail locations increased by 5-10% over the course of 6 months.

In terms of automation of store format and assortment management, four key tools were developed to manage and optimize store space:

1. Location Monitoring Dashboard. This tool allows for the assessment of a store’s potential in terms of revenue, margin, or other key performance indicators of choice. It can then determine the key success factors and risks for the location, as well as compare performance metrics with those of other similar locations.

2. Heat Map of Potential Stores. This tool enables the organization to create visual results on a map, highlighting areas with the highest potential. It also provides the ability to view detailed information (e.g., demographics, competition, store density) through a "double-click" function.

3. Optimization of New Store Locations. This feature determines the optimal number of stores and their locations, evaluates the return on investment (ROI) for each location, and allows scenario analysis based on business assumptions.

4. Setting Sales Targets, Managing Performance, and Choosing Store Closure Options. This includes a comparative analysis of the store’s key performance indicators with other similar stores in similar locations (see Fig.4).

Fig.3. An example of a map.

Initial data used:

- Residential buildings with aggregated data on the number of apartments and population breakdown by age;

Based on the data, an information panel was developed, providing details on store characteristics, key performance indicators, and its driving forces (Fig.4).

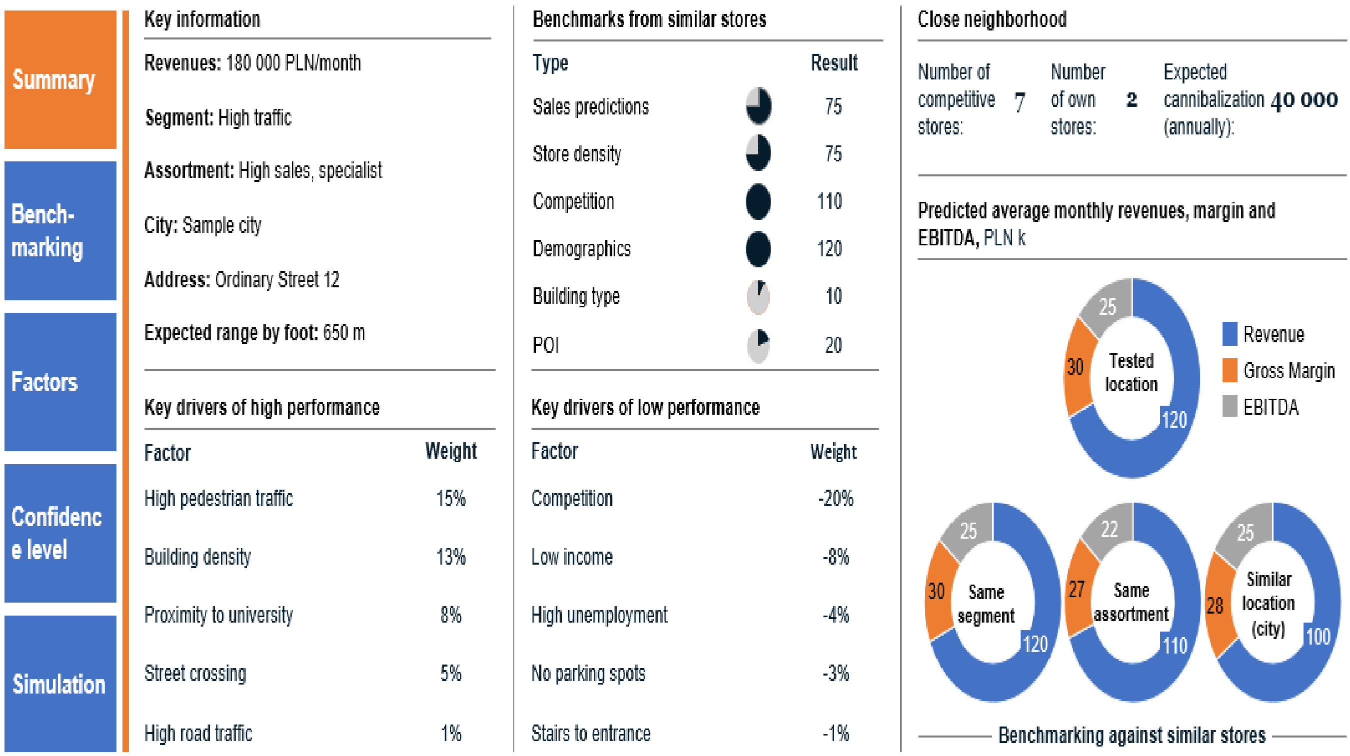

Fig. 4. Information panel.

Next, let's consider the process of creating a heatmap based on the available data. The map displays address points that are further analyzed. Each location is scored, with the monthly revenue forecast based on geospatial analysis using:

- Demographic data (e.g., population density, income, age, education);

- Competitor data (e.g., proximity to competing stores);

- Pedestrian traffic data;

- POI (proximity to traffic-generating objects like offices or tourist attractions);

- Weather data (e.g., deviations from typical weather conditions based on data from 100 measurement points);

- Performance data (e.g., store size, quality audit results).

The results obtained from the model are interpolated to identify areas with similar potential (Fig. 5).

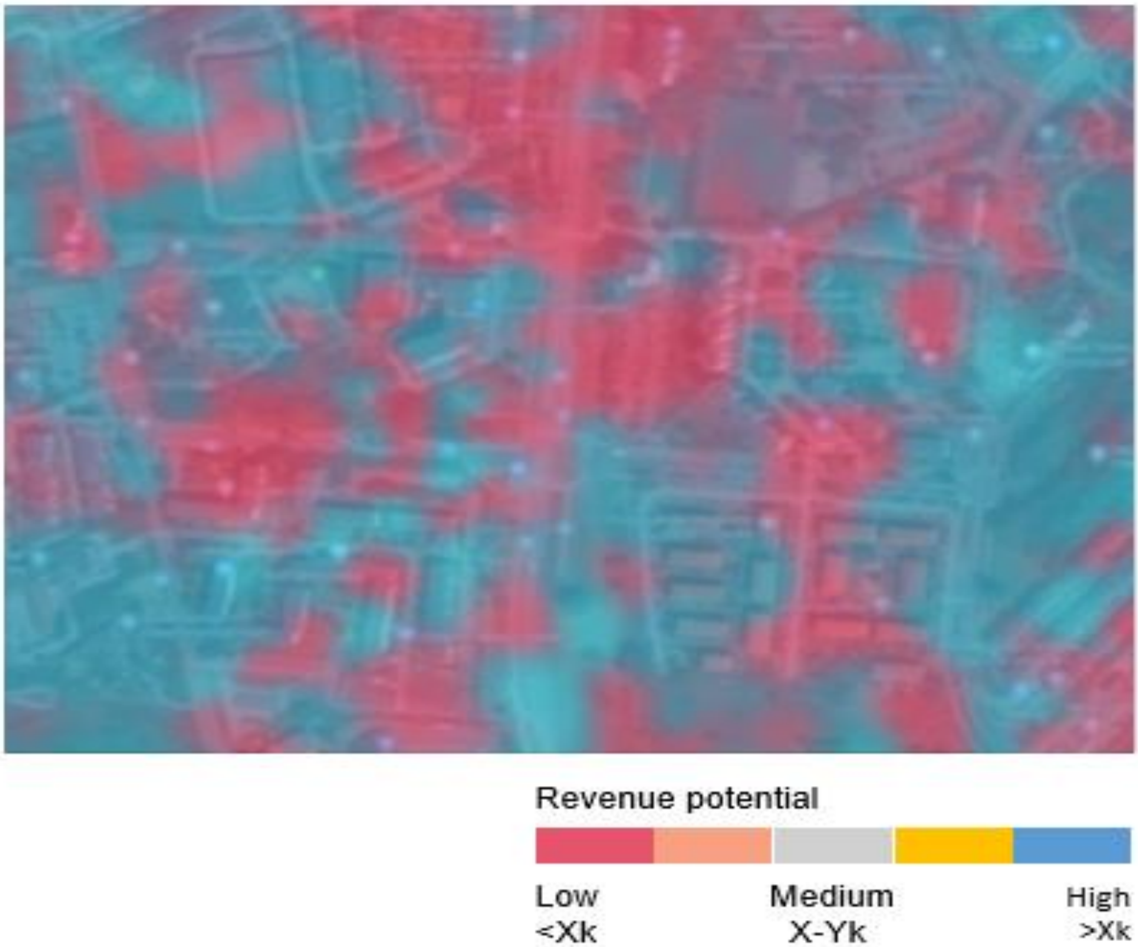

Fig. 5. Heatmap of sample residential estate.

Evaluation of return on investment (ROI) for a new store. A separate EBITDA assessment model was developed for new stores based on:

- Expected revenue;

- Assortment and store segment affecting the margin;

- Operating costs associated with the type of location.

For each location, the model forecasts:

- Gross profit;

- Logistics costs;

- Rent expenses;

- Personnel costs;

- Other expenses.

Based on the above, the payback period was also calculated (Fig. 6).

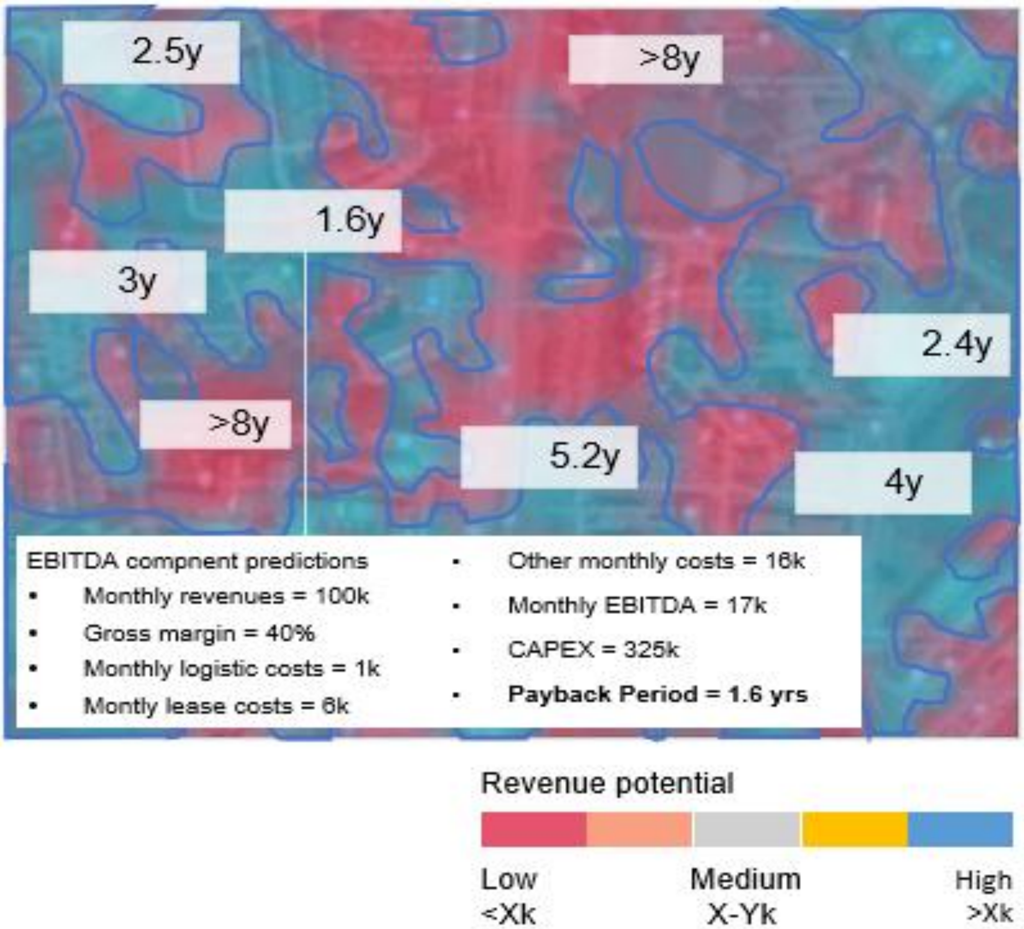

Fig. 6. Heatmap of sample residential estate.

Next, we look at geocoding customers and forecasting probabilities using deep analytics models with individual characteristics (residence/workplace, digitalization, grocery penetration, etc.). The optimization process simultaneously compares the possibility of relocation to multiple locations, which significantly complicates the calculations. This method can be used to refine customer flow for the final footprint and assess individual actions.

Advantages:

- Universal function for calculating churn/flow;

- Depends on the number of neighboring branches and distance;

- Accounts for actual city/district-level distribution to reflect travel tendencies;

- Has a probabilistic interpretation.

The main disadvantage of this method is that actual customer distribution and flow per specific branch may vary significantly.

Following changes in the system, cannibalization variables from the current network are updated, and the potential in all areas within the impact zone is recalculated. Effective data processing tools are required to use this method. It can be used to refine customer potential for the final footprint and evaluate specific actions. By using dynamic updates of geo-potential maps, it is possible to account for non-linear relationships and the actual distribution of traffic generators on the map. However, this method requires significant computational power, and, considering the specifics of the training set, the results may not align with economic logic in edge scenarios.

CONCLUSION

Thus, the application of unsupervised machine learning and vector models in the design and optimization of retail channels for telecommunications services opens new horizons for improving business efficiency. These methods allow for the discovery of hidden patterns in large data sets, enhance the accuracy of demand forecasting, and optimize logistics processes. Despite the challenges associated with implementing these technologies, their use provides significant competitive advantages, improving customer experience and the company's operational performance. In the future, further research in this area may contribute to the development of more advanced models and methods capable of even more effectively addressing the optimization tasks of sales channels in the telecommunications industry.

REFERENCES

1. Zaadnoordeik L., Besold T. R., Cusack R. Infant learning lessons for unattended machine learning //Nature Machine Intelligence. – 2022. – Vol. 4. – No. 6. – pp. 510-520.
2. Allogani M. et al. A systematic review of supervised and unsupervised machine learning algorithms for data science //Supervised and unsupervised learning for data science. - 2020. – pp. 3-21.
3. Fadokun D. O., Oshilike I. B., Onyekonwu M. O. An approach to machine learning with and without control in forecasting facies // Annual International Conference and Exhibition SPE in Nigeria. – SPE, 2020. – Number D013S014R007.
4. Ma S., Fildes R. Forecasting third-party mobile payments, taking into account the consequences for forecasting the flow of customers //International Journal of Forecasting. – 2020. – vol. 36. – No. 3. – pp. 739-760.
5. Guven I., Shimshir F. Forecasting demand using color parameters in the retail clothing industry using artificial neural network (ANN) and support vector machine (SVM) methods //Computers and Industrial Engineering. - 2020. – Vol. 147. – p. 106678.
6. Jansen S. Machine Learning for Algorithmic Trading: Predictive models for extracting signals from market and alternative data for systematic trading strategies using Python. – Packt Publishing Ltd, 2020.
7. Ho L., Getals P. Application of machine learning in river research: trends, opportunities and challenges //Methods in ecology and evolution. – 2022. – vol. 13. – No. 11. – pp. 2603-2621.