Data Exchange Platform for Digital Economy Applications

Oleg Surnin
Data analysis department
IPSI SEC "Open Code"
Samara, Russia
surnin@o-code.ru

Pavel Sitnikov
Faculty of infocommunication
ITMO University
Saint-Petersburg, Russia
sitnikov@o-code.ru

Anastasia Khorina
Faculty of infocommunication
ITMO University
Saint-Petersburg, Russia
anastasiakhorina@mail.ru

Anton Ivaschenko
Computer science department
Samara State Technical University
Samara, Russia
anton.ivashenko@gmail.com

Anastasia Stolbova
Information systems and technologies department
Samara National Research University
Samara, Russia
anastasiya.stolbova@bk.ru

Nataly Ilyasova
Image Processing Systems Institute of RAS - Branch of the FSRC "Crystallography and Photonics" RAS
Samara, Russia
ilyasova.nata@gmail.com

***Abstract*—This paper describes an experience of the Data exchange software platform practical use supporting the modern trends of Digital Economy. The platform was initially designed for the suppliers and customers of data sources providing the up-to-date technologies of big data processing as an online service. The platform is also open for software developers to upload new algorithms and technologies in order to help them to find new areas of application. There is presented architecture and its software implementation for an intermediary online platform capable of collecting, processing and analysis of various datasets. Modern companies being the members of digital economy can use this platform to process their data and produce business analytics. They can become both suppliers and providers of data, as well as develop and upload new customized algorithms. First results were achieved in the area of retail and social media analysis.**

Keywords—data exchange, digital economy, big data

I. INTRODUCTION

Data exchange platform is a new concept to organize an efficient cooperation between the providers and customers of various data sets and algorithms based on implementing the software solution of open service provider [1 – 2]. Considering the results of open service software providers' practical use [3], there was developed and probated new software architecture capable of solving the actual problems of business analytics in the different spheres of digital economy.

Such a solution is presented below with an illustration of Data exchange platform application in network retail. Data exchange platform implements the modern technologies of big data analysis and capable of providing business analytics and decision-making support in real time. The proposed solution architecture and case study intended for processing network retail data, predicting the processes, managing the placement of big data, and planning the computing load balancing.

II. PROBLEM DOMAIN OVERVIEW

Retail is a popular way to organize distributive trades. Modern retail industry becomes a promising area of development of network organizations that provide distributed services for a large amount of consumers and therefore processing the concerned flows of big data.

The logic of this information processing is mainly influenced by the processes of retail service. The following areas of retail are distinguished:

- street retail – organization of retail trade in the most visited places: on pedestrian streets, ground floors of buildings;
- non-food retail – organization of trade in non-food products, which in grocery stores, as a rule, are called related. Such products include clothing, cosmetics, household chemicals, stationery and other categories of goods;
- food retail – organization of food trade. This category of goods is the most demanded and relates to everyday goods;
- network retail – organization with several stores of the same chain united by one concept;
- electronic retail – organization of trade through the Internet;
- cellular retail – organization of trade between mobile operators.

Retail industry gains significant benefits of being adaptive to customer changing demands. Understanding the trends and prediction of changes helps reducing costs and increasing competitiveness. Application of modern technologies for data collecting and processing can solve this problem.

Retailers use the following mechanisms to successfully organize business processes that form the main requirements for a data processing toolset:

- calculation of the most optimal locations for placement of outlets;
- use of modern commercial equipment;
- work with categories of clients;
- work with methods of attraction;
- development of self-service and reduction of personnel due to this;
- optimization of logistics, work with wholesale suppliers;
- automation of all stages of trade.

Depending on the specifics of retail problem domain and other factors, a retail development strategy is determined. The strategy operates with critical factors, by solving the

Copyright © 2020 for this paper by its authors.
Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0)

CEUR-WS.org/Vol-2667/paper53.pdf

problems of determining the selling price of goods sold, managing the assortment of outlets, and determining their location.

A comparison of the most frequently used business intelligence systems in the retail sector for a number of the most important criteria is shown in Table 1.

TABLE I. PLATFORMS COMPARATIVE ANALYSIS

Feature	Qlik View	Klip folio	Tab leau	Power BI	Data ex change
SQL and software develoment skills	+	-	-	-	+
Own programming language	+	-	-	-	+
Connectivity to heterogeneous data sources	+	+	+	+	+
Data storage	+	+	+	+	-
Data aggregation	+	-			+
A variety of dashboards for visualization	+	+	+	+	+
Graphical programming language	-	-	-	-	+
Intelligent algorithms and data processing	-	-	-	-	+
Specialization in retail analytics	-	-	-	-	+
Possibility of monetization of data, analytics and algorithms	-	-	-	-	+

Currently, electronic retail is popular and its popularity is growing according to the main trends of digital economy development. Electronic retail includes trade in both food and non-food products, which leads to serious competition among retailers. Price is the main factor that allows companies to stand out among their competitors and attract buyers, so pricing strategies should be the most flexible.

There are a variety of approaches to pricing:

- personalized approach to customers with the formation of individual offers: this approach is based on the analysis of the consumer behavior of each specific buyer, the determination of his needs, financial capabilities and, ultimately, the formation of individual price offers for him;
- promotion management: the approach is aimed at the formation of special price offers for goods or groups of goods according to certain criteria in order to attract customers. In applying this approach, it is important not to create a reputation as a discounter, which can negatively affect the perception of the retailer by customers, so promotions should be temporary;
- psychological techniques that affect the perception of price. This approach is common for the retail industry as a whole and is quite popular, due to the ease and low cost of its application. For example, to sell an expensive product, you need to put it next to an even more expensive product. Another fact that has been studied is that the customer has a high probability that he or she will not have a single price, so a small difference in price should be left for similar products. Psychological techniques include managing delivery prices. It is very often necessary to arrange delivery. Modern research shows that most shoppers leave the store basket. The offer of free delivery or promotional goods when making delivery positively affects the customer;
- predictive pricing is a fairly effective approach, the main task of which is to determine how price affects demand. The approach allows you to model changes in demand depending on the price, taking into account various factors, including other methods involved in pricing: psychological pricing, the influence of competitors, behavioral models and categories of current retailer customers, current trends (in the fashion industry or global trends in lifestyle changes).

III. STATE OF THE ART

The results of [3] show that retailers are moving to more innovative strategies to offer modern consumer solutions based on technological advances. The high level of property rights, due to the enormous number of patents, forces retailers to invest more in the acquisition of patented technologies to achieve advantages over competitors or to introduce new management methods.

The need to analyze retail data in a highly competitive environment is shown in [4]. Advances in machine learning and big data lead to the use of data analytics management systems in many organizations and industries.

As a rule, large organizations can devote more resources to this, and the software used is more suitable for large enterprises. However, the growing business pressure is forcing small and medium-sized enterprises to implement data analytics, which is new to them and leads to a number of problems considered in [5].

Social networks have become a part of life for most people around the world. Retailers (and not only) are actively using them to share information about their products with customers. As a result of the growth of social networks, the need for monitoring, data mining and analysis is increasing. So, in determining the tasks and opportunities of network retail, the issues of integration with social networks, determining a development strategy and studying the life cycle of clients are relevant [6, 7]. The study of customer behavior models requires taking into account completely diverse data and improving their analysis.

In [8], the questions of creating a body of knowledge are considered, the concept and key methods of data analysis are studied in the field of retail network, the creation of data exchange methods [9], such as shopping basket analysis [10] and sales analysis in general using business intelligence tools.

Analysis of factors of customer satisfaction and loyalty, the image of retailers and the relationships between them lead to the emergence of models for creating a satisfactory experience for consumers, which is a priority for retailers

[11]. The influence of fluctuations in demand and purchasing power on macroeconomic regulation and state control are considered in [12].

Research on the effects of loyalty programs on the decision to purchase goods at different periods of time is relevant, since an important factor for multichannel grocery retailers is which promotion strategy to choose across all channels. The tendency of product buyers to continue to visit an offline store after they start buying in the online store of the network means that promotions in one channel can have a significant negative impact on the behavior of customers in another channel, especially if promotions differ by channel [13].

To solve the above mentioned problems of network retail there can be used a variety of data sources. For example, it is possible to analyze data obtained in the following areas:

1. Acquiring data which includes information about the commission of the acquiring bank, categories of outlets, payment card options, which allows you to determine the dependence of the availability of sales outlets on the availability of POS terminals;

2. Retailer data, which include:

- information on loyalty programs and promotions to determine the dependence of demand on the availability of various special offers for goods;
- information about regular customers and users of loyalty programs to analyze the dependence of the purchased goods on the categories of customers, to form the most advantageous offers for both the buyer and the retailer. In addition, it is possible to evaluate the customers in terms of how much money they spend and how often they make purchases in retailers' stores, which will determine the significance of each customer for the network and more accurately formulate personal offers;
- date of purchase and time of purchase to analyze the dynamics of demand for certain products depending on seasonality and time of day. In addition, it is necessary to take into account various holidays and ongoing events in the analyzed region;
- the list of goods on the check allows you to analyze the most frequently purchased goods together;
- the price of the items listed in the check and the total purchase amount allow the analysis of the turnover of the outlet;
- the trajectory of buyers allows you to determine the order of purchases at retail outlets and to identify the optimal location of goods in the store;

3. Social media that help identification of fashion trends and identify the dependence of purchases as well as the most popular brands and products;

4. Geo data that can be used to determine the dependence of consumer activity, demand for goods, payment methods depending on the territorial location of the outlet and the nearby infrastructure, such as the area (sleeping, tourist, business centers and others), categories of users visiting or living in the area.

According to the commission of the Code of good practices, it is known that retailers do not want to share sales data of individual outlets with suppliers, which leads to an inaccurate assessment of their activities, loss of productivity, and, consequently, loss of income.

IV. Solution Architecture

Having considered the tasks and problems of network retail, strategies for solving them, we can conclude that the creation of a data exchange platform is a necessary step for the business development, and its intellectualization is an integral part of the process.

Intellectualization includes the introduction of the following technologies into the data exchange:

1. Face recognition. Biometric assessment systems can optimize salary costs, allow you to calculate the costs of non-staff employees and minimize the risks of dishonest actions and counteractions. The introduction of face recognition can also significantly increase understanding of the characteristics of the average buyer. Indeed, many algorithms allow you to evaluate gender, age, race, and thereby form a portrait of the buyer. To solve such problems, a camera is used, aimed close-up at incoming visitors, with a face detection function. Additionally, such algorithms in combination with the display of advertising can give an understanding of the effectiveness of advertising on visitors.

2. Creating smart baskets. Built-in cameras that can recognize and scan products, sensors that detect the detection of an object in the basket, as well as scales that allow you to get rid of additional weighing of fruits and vegetables. On the trolley screen, the buyer will be able to see all the products taken, as well as their total amount. The data that carts collect (on routes through stores, on the frequency of purchases of goods from certain shelves, where customers are, etc.) would help company partners optimize their stores.

3. Analysis of social networks. Analysis of data from social networks will allow you to analyze customer complaints in real time and track user requests and wishes. Location and shopping data will allow you to compile the most complete portrait of a specific retailer buyer.

4. Internet of things. When it comes to retail, the IoT infrastructure includes RFID tags, infrared traffic meters in stores, satellite and Wi-Fi tracking systems, digital signatures, kiosks or even mobile devices of the customers themselves. The Internet of things makes it possible to implement predictive maintenance of equipment, transportation, keeping a warehouse on the basis of demand, tracking the activity of buyers, creating a smart store, behavioral analytics, personalized marketing and real-time advertising based on location and purchase history.

5. Adaptive Acquiring. The sources of such a data exchange are intelligent systems. Data processing can give conflicting results in different contexts and in the absence of data, but the results obtained should provide decision-making support. Technologies without data will not give the desired result, as well as data without applying the necessary technologies to them. So, the data exchange can be presented as a platform for interfacing digital intelligent services.

The proposed software solution for a data exchange platform is presented in Fig. 1. It was implemented using Java (IntelliJ IDEA 15.0.3 Integrated Software Development

Environment (Community Edition)) and supports JavaScript, CoffeeScript, HTML / XHTML / XAML, CSS / SASS / LESS, XML / XSL / XPath, YAML, ActionScript / MXML, Python, Ruby, Haxe, Groovy, Scala, SQL, PHP, Kotlin, Clojure, C, C ++.

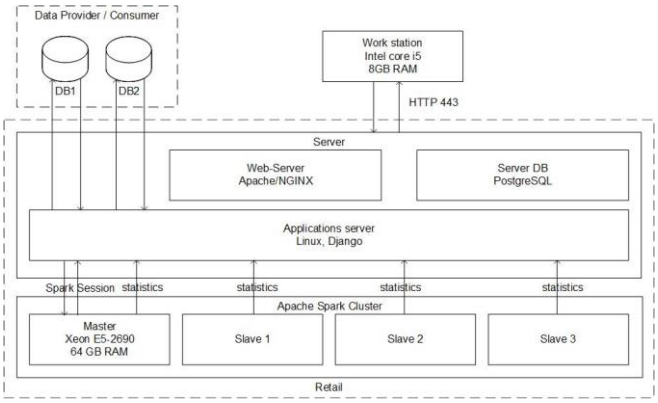

Fig. 1. Software architecture.

Additional libraries and frameworks include Apache Hadoop, Apache Spark, and Django. Apache Hadoop (HDFS) is a file system designed to store large files, block-by-block distributed between nodes of a computing cluster. Apache Spark is an open source framework for implementing distributed processing of unstructured and weakly structured data, which is part of the Hadoop project ecosystem.

The developed data exchange platform provides the functionality for pre-processing and analysis of network retail data. As a part of data pre-processing the system implements the methods of data structuring, omission processing, and data dimension reduction, trend highlighting, and correlation analysis.

To solve the problems of network retail analysis, there is a possibility to use the following:

- Apriori, FPG (Frequent-Pattern Tree) methods help solving the problem of analysing a shopping cart;
- Linear regression methods are used to solve the forecasting problem;
- The Mean method is used to calculate average values;
- K_means method is used for various kinds of clustering.

Besides, a wide range of methods such as ARIMA, FB Prophet and others are used to analyse time series.

V. Implementation and Tests

The following example illustrates the results of Data exchange platform development. We create a project for personal offers on selected categories of goods formed by a retailer for users of a social network with loyalty cards. Forming this request is useful if you need to increase sales for certain categories of goods available to the retailer.

The result of this query is a list of users who are interested in purchasing goods for selected groups of goods with an indication of the group. By analyzing the results obtained, the retailer can formulate personal offers for the groups of goods chosen by him for the most suitable customers.

In the considered example, users were searched for the following product groups: 3 (energy granola bars), 4 (instant foods), 5 (marinades meat preparation). As can be seen from Figure 2, for the buyer with the number 141848112, you can create an offer for goods from group 3 (energy granola bars), and for the buyer with the number 78310286 – 4 (instant foods).

Fig. 2. Project Results.

Another popular task in the field of network retail analysis is the sales forecast. The following fields were selected as the initial data of the problem to be solved: store identifier, purchase date, purchase amount.

To solve the problem, there were combined the basic methods powered by the developed data exchange platform:

- python_basic_filter – the method allows you to select sales data for one store from many;
- python_basic_resample – the method allows you to aggregate data by day, summing them up;
- python_timeseries_holt_winters – the method allows forecasting using the Holt-Winters additive model (triple exponential smoothing).
- python_plotly_forecast – a method for plotting.

The ability to combine methods makes it possible to choose the most suitable analysis algorithm, taking into account the specifics of the source data. The project drawn up in the query designer is shown in Fig. 3.

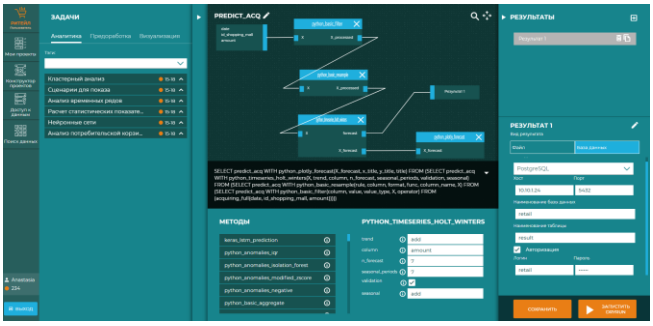

Fig. 3. Predictive analysis project.

The result of this project is the forecasted revenue values of the selected store, taking into account the seasonality period with a sales horizon.

The proposed approach opens new perspectives for online services providers and buyers cooperation in common information space. According to the current trends of economy digital transformation most services migrate to web platforms transporting all negotiations between the business parties to virtual reality.

Therefore social media become a substantial part of business relations. The resulting effect forms the economy of ultra-low expenditures. On the one hand, the members on such relations can swiftly change their mind by getting better options, producing the contract agreements that are easy to enter and easy to leave. On the other hand, all the games are being fixed, and the members agree to play open.

Under these conditions a new class of intermediary open service platforms starts playing the general role. Using the existing IT infrastructure they allow building a new virtual world powered by business analytics. In this sphere decision-making can no longer be done by humans themselves. All the decisions need a support from data analysis and machine learning, which makes the access to them even more important than economy resources available in real life.

VI. Conclusion

The developed solution based on the data exchange platform allows predicting the processes of network retail manage the placement of big data and plan the computing load based on the info logical model, which will increase the sales efficiency.

Next steps are related to extending the area of intermediary open service application including the problem domains of social analysis and services automation. Positive results were also achieved in banking acquiring, which can be extended in business sphere making the data processing an effective tool of economy digital transformation.

Main research results include the following deliverables: big data management methodology (O. Surnin), and its implementation by Open code platform architecture (P. Sitnikov), analysis of its perspective for digital economy (A. Ivaschenko) based on implementation of data exchange model (A. Stolbova) and algorithms of semantic (A. Khorina) and statistics (N. Ilyasova) analysis.

Acknowledgment

This work was financially supported by the Russian Foundation for Basic Research under grant # 19-29-01135 and by the Ministry of Science and Higher Education within the State assignment to the FSRC “Crystallography and Photonics” RAS.

References

[1] O.L. Surnin, P.V. Sitnikov, A.V. Ivaschenko, N.Yu. Ilyasova and S.B. Popov, “Big Data incorporation based on open services provider for distributed enterprises,” CEUR Workshop Proceedings, vol. 190, pp. 42-47, 2017.

[2] O.L. Surnin, P.V. Sitnikov, A.A. Khorina, A.V. Ivaschenko, A.A. Stolbova and N.Yu. Ilyasova, “Industrial application of big data services in digital economy,” CEUR Workshop Proceedings, vol. 2416, pp. 409-416, 2019.

[3] E. Pantano, “Does innovation-orientation lead to retail industry growth? Empirical evidence from patent analysis,” Journal of Retailing and Consumer Services, vol. 34, pp. 88-94, 2017.

[4] C. Kahraman, “Efficiency analysis in retail sector: implementation of data envelopment analysis in a local supermarket chain,” Cham: The International Symposium for Production Research, Springer, pp. 884-897, 2018.

[5] M. Mirzaei, “Investigating challenges to SME deployment of operational business intelligence: a case study in the New Zealand retail sector,” Proceedings of the 12th IEEE/ACM International Conference on Utility and Cloud Computing Companion, pp. 139-142, 2019.

[6] O.G. Ayodeji and V. Kumar, “Social media analytics: a tool for the success of online retail industry,” International Journal of Services Operations and Informatics, vol. 10, no. 1, pp. 79-95, 2019.

[7] I.A. Rycarev, D.V. Kirsh and A.V. Kupriyanov, “Clustering of media content from social networks using BigData technology,” Computer Optics, vol. 42, no. 5, pp. 921-927, 2018. DOI: 10.18287/2412-6179-2018-42-5-921-927.

[8] F. Castelo-Branco, “Business intelligence and data mining to support sales in retail,” Singapore: Marketing and Smart Technologies, Springerpp, pp. 406-419, 2020.

[9] Y. Chen, “FaDe: a blockchain-based fair data exchange scheme for Big Data sharing,” Future Internet, vol. 11, no. 11, pp. 225, 2019.

[10] M. Kaur and S. Kang, “Market basket analysis: identify the changing trends of market data using association rule mining,” Procedia computer science, vol. 85, pp. 78-85, 2016.

[11] C.M. Veloso, “The loyalty and satisfaction determinants: a factor analysis applied to the south and insular Portuguese traditional retail,” 11th Annual Conference of the EuroMed Academy of Business: Research Advancements in National and Global Business Theory and Practice, EuroMed Press, pp. 1380-1394, 2018.

[12] S. Shen, “An empirical analysis of the total retail sales of consumer goods by using time series model,” Journal of Mathematical Finance, vol. 9, no. 2, pp. 175, 2019.

[13] E. Breugelmans and K. Campo, “Cross-channel effects of price promotions: an empirical analysis of the multi-channel grocery retail sector,” Journal of Retailing,vol. 92, no. 3, pp. 333-351, 2016.