Vasil Marchev
Angel Marchev, Jr

METHODOLOGICAL IMPLEMENTATION OF CRISP-DM IN FINTECH SERVICES

Vasil Marchev

Chief assist. Prof., PhD

Department of Management/ Faculty of Management and Administration, UNWE, Sofia, Bulgaria

e-mail: vmarchev@unwe.bg

ORCID: https://orcid.org/0000-0001-9412-2831

Angel Marchev, Jr

Professor, PhD

Department of Management/ Faculty of Management and Administration, UNWE, Sofia, Bulgaria

e-mail: angel.marchev@unwe.bg

ORCID: https://orcid.org/0000-0002-5090-3123

Published First Online: 27.06.2025

Pages: 58-71

DOI: https://doi.org/10.37075/JOMSA.2025.1.06

ABSTRACT

The financial markets are undergoing a significant transformation due to the changing behavior of individual investors. This shift is moving away from traditional investment options towards more complex financial products. Advancements in technology and a greater understanding of investments have contributed to this change by diminishing trust in conventional financial institutions. The integration of big data analytics within FinTech is essential for understanding client behavior and personalizing services to meet individual needs. By employing methodologies such as the Cross Industry Standard Process for Data Mining (CRISP-DM), FinTech firms can systematically extract insights from data, ultimately enhancing customer satisfaction through tailored investment strategies. The CRISP-DM framework outlines critical stages for effectively managing the complexities of data-driven decision-making in the financial sector.

This paper discusses the methodological implementation of CRISP-DM in the FinTech industry, highlighting its role in optimizing service delivery and improving customer engagement through data analysis and strategic modeling.

KEYWORDS: fintech industry, CRISP-DM, individualization, data analyses

JEL: M10, G40, C81

INTRODUCTION

The revolutionization of financial markets finds its roots in the dynamic change of individual investors. A new niche based on investment opportunities is emerging due to the changed customer attitude. There is a migration from traditional investment opportunities to complex financial products.

The return of the individual investor is rooted in the change that has occurred in the investment markets, in terms of the development of technology companies. There are also changes in traditional financial structures and the savings products they offer. Finally, a positive trend is being established in the growth of the volume of knowledge in the field of investments (Marchev & Marchev, 2023c). The factors listed above are leading to revolutionary trends in a conservative market, such as the market for financial services and investments.

The revolution in financial markets is primarily driven by the evolving behavior of individual investors. A new niche focusing on diverse investment opportunities is emerging due to shifts in customer attitudes. There is a transition from traditional investment options to more complex financial products.

1. Essentials of the FinTech industry

The FinTech sector represents a modern approach to delivering financial services, leveraging advancements in technology that have appeared over recent decades. Companies within this space are responding to the rapid evolution of technology by positioning themselves as alternatives to traditional financial institutions, offering services that are not only faster but also more adaptable to consumer needs. The aftermath of the 2008 financial crisis, marked by a significant erosion of trust in conventional financial entities, has catalyzed the growth of the fintech industry.

FinTech firms focus on enhancing customer service experiences by providing financial solutions that are responsive to the changing dynamics of the global market. This includes innovative strategies for managing customer expectations, such as implementing advanced monitoring systems aimed at maximizing overall satisfaction. The industry's goal is to create a more efficient and personalized financial service landscape, which is increasingly essential in a world where consumer demands are continually evolving.

The development of financial systems is driven by the rapid and dynamic evolution of new technologies. According to Chaudhry et al. (2022), the short timeframe within which the tangible impact of financial technology evolution is felt significantly impacts the performance of financial activities and transactions with customers.

Fintech companies focus their efforts on part of the financial services available in the portfolio of traditional representatives of the banking system. The main point of distinction between these service providers is rooted in the way they are provided. The advantage that FinTech has over traditional financial institutions is the flexibility of its services in meeting customer needs. They provide new channels for delivering the desired product while improving the process, both in terms of financial aspects and in terms of speed and convenience.

Fintech companies are distinguished in several key domains where they offer more competitive products compared to traditional commercial banks. These areas include money transfer services, digital wallets, lending solutions for small and medium-sized enterprises (SMEs), asset management, and various other financial services.

The evolution of the entire palette of financial services from conventional to electronic form is realized through the development of new technologies, and in particular, the opportunities provided by the development of the Internet.

Financial agents in the economy aspire to satisfy customer needs more fully, which is directly related to improving the quality of services offered. In line with this, a new range of modern products is being developed to further increase customer satisfaction.

FinTech based on big data analyses

In recent decades, the impact of big data on society and science has become undeniable. The financial system, as part of the broader economic landscape, cannot escape the rapid advancement of large volumes of information. The influence of this data on the development of financial systems has been well established. The topic was explored by Mhlanga (2024).

The nature of extensive data and the analytical methods associated with it enable financial institutions to gain a comprehensive understanding of their clients. As a result, they can offer a new type of service tailored to meet the individual desires and needs of their customers.

According to Costa and McCrae (1989), "personality traits are continuous and consistent characteristic reactions of individuals when interacting with different circumstances." These personality aspects are key drivers of individual behavior; they offer insights into expected responses in various situations and how individuals interact with their environments.

Beyond personal characteristics, general qualities also significantly influence behavior. These qualities are a set of behavioral norms typical of most people, although individuals may possess them to varying degrees. They serve as benchmarks for behavioral analysis (Allport, 1961).

By considering both individual and general characteristics, we can create an individualized approach to developing and managing investment portfolios. This approach considers a range of previously identified traits deemed important in selecting investment opportunities (Marchev & Marchev, 2023c). The personalized strategy offers a variety of investment options tailored to the preferences, desires, and needs of individuals.

The abundance of information in financial services needs a structured approach to understanding, analyzing, and extracting insights from consumer behavior. The CRISP-DM methodology provides an appropriate sequence of actions for this purpose.

2. Cross Industry Standard Process for Data Mining (CRISP-DM)

The industry standard process method for extracting knowledge from data - Cross Industry Standard Process for Data Mining (CRISP DM), is a set of interconnected and sequential steps for understanding and discovering the needs of the business under consideration, reviewing and analyzing information, building a model, and validating the process (Leaper, 2009; Schröer et al, 2021).

The method under consideration is an open standard that is available to all interested parties. It was developed by a consortium of over two hundred interested organizations with funding from the European Union. The model is focused on the data mining process but assumes the necessary flexibility to respond to a wide range of analytical styles (Brown, 2015). In 1996, Daimler-Benz, Integral Solutions Ltd. (ISL), NCR, and OHRA began the development of a single standard for extracting information from data. A dynamic development in the direction of targeted data mining followed, and a consortium was formed to improve the new approach. In terms of design and implementation, the method is neutral concerning the application industry. To fulfil this task, representatives of different industries and businesses were integrated into the project. The method integrates an interdisciplinary approach aimed at standardizing the processes under consideration. Colin Shearer writes more on the subject (Shearer, 2000).

CRISP-DM is based on a set of initial attempts to standardize a methodology for extracting knowledge from data (Reinartz, Wirth, 1995; Adriaans, Zantinge, 1996; Brachman, Anand, 1996).

The process plays a vital role in comprehending and overseeing the interactions within this intricate process. The adoption of a standardized process model in the market offers numerous advantages. It can act as a shared reference framework for discussions surrounding data mining, enhancing the collective understanding of key data mining challenges among all stakeholders, particularly from the customer perspective (Wirth, Hipp, 2000). A single standard is established to improve understanding by unifying criteria and framing the overall process

For the purposes of this paper, a detailed overview of the processes has been made, following the logical sequence of the industrial standardized process method for extracting knowledge from data. The method under consideration is used in the research as the main toolkit for processing information arrays.

Business understanding

The first phase of the method is called Business Understanding. This subsystem includes several components that define as their main task the correct and adequate identification of the business situation under consideration.

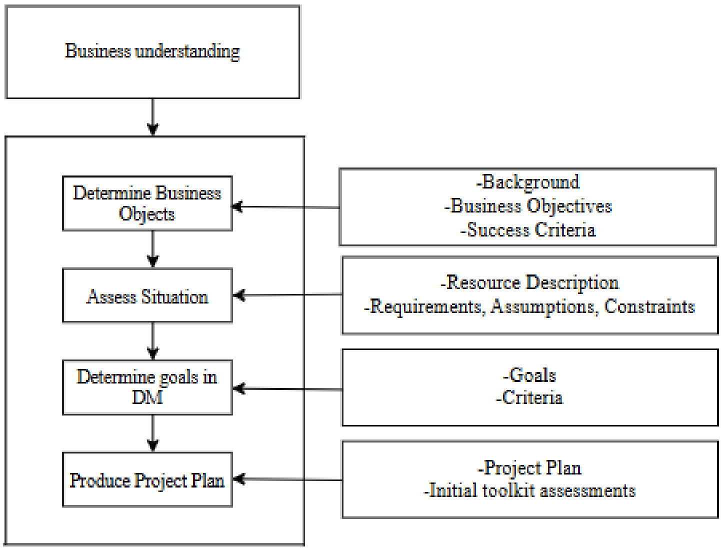
It is essential to define the main business objectives that are set before the model integrating the CRISP-DM method. Business goals are generally in correlation with the main aspects of the business/organization under consideration. They should be in harmony with both the historical achievements of the company and its current and future development plans.

An essential moment in understanding the business situation, in addition to accurate and appropriate goal setting, is establishing the main criteria for success. The specified criteria can be determined as certain desired states of a series of indicators for which a specific measure is available.

In work, it is necessary to carry out a complete analysis of the current situation. The correct and adequate assessment of the positive and negative features has a significant role in the adequate application of the method under consideration. It is necessary to conduct a thorough review of the available resources to set up the main requirements, assumptions, and limitations from the point of view of the algorithm of action. Establishing the main risks and a series of rules in case of unforeseen circumstances is essential for adequately calculating the expected benefits and costs associated with the project.

Extracting information from data is a complex, labor-intensive, and time-consuming process. However, if correct and situation-appropriate goals related to this process are generated, it can have significant benefits for the project's development. For this purpose, it is necessary to set proper criteria for the relevance of the tasks set before the algorithm.

Figure 1 Business Understanding Subsystem Structure



Source - own

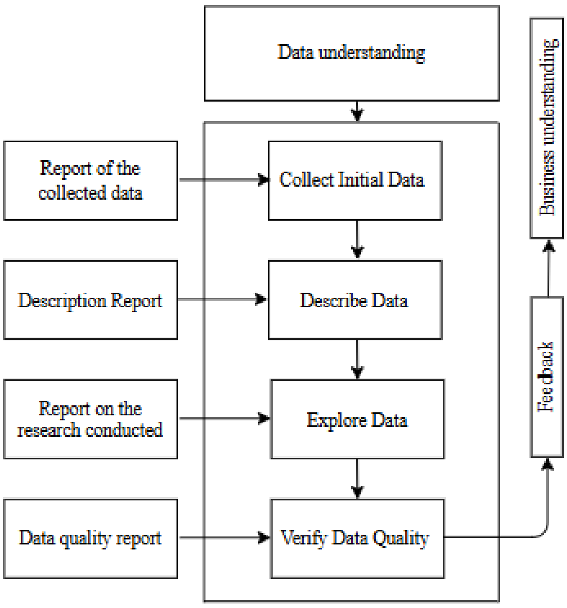
The phase of the process of understanding the business environment passes through the preparation of a project plan. From the point of view of the method it is necessary to determine the toolkit of work. It includes the requisite tools for conducting the project, software security, technical training of the participants and the technical aspects of the goals set above. Initial assessments are made of the available toolkit and set of techniques.

Data understanding

Next step in the considered standardized data extraction method is related to the process of insight into the essence of the available information arrays (Fig. 2). The observed phase starts with actions to provide the model with the necessary amount of information. Data related to the studied system are extracted using a set of algorithms. This is a process in which it is required to generate an array that is maximally consistent with the goals and criteria set in the first phase of the method.

Following the data extraction step, the next phase involves a thorough examination and description of the available data sets, focusing on their specifics and characteristics. The primary goal of this subphase is to gain an in-depth understanding of the collected data sets and their alignment with the established goals. After preparing a comprehensive description of the data, proceed to the detailed study and analysis.

Figure 2 Subsystem structure Understanding data



Source - own

Data analysis is the stage where the alignment of available information with established criteria and goals is evaluated. Monitoring this alignment can reveal challenges in determining project affiliation. If discrepancies are found, the method allows for revisiting the initial phase to revise the set goals. Additionally, this phase offers the opportunity to identify shortcomings in understanding the nature of the business. The primary tasks during the analysis phase involve conducting research and defining the characteristics of the data.

If the results obtained satisfy the set goals, it proceeds to verification of the data sets in correspondence with the set criteria for compliance. The stage concludes with the preparation of a report on the quantity and quality of the available data sets.

Data preparation

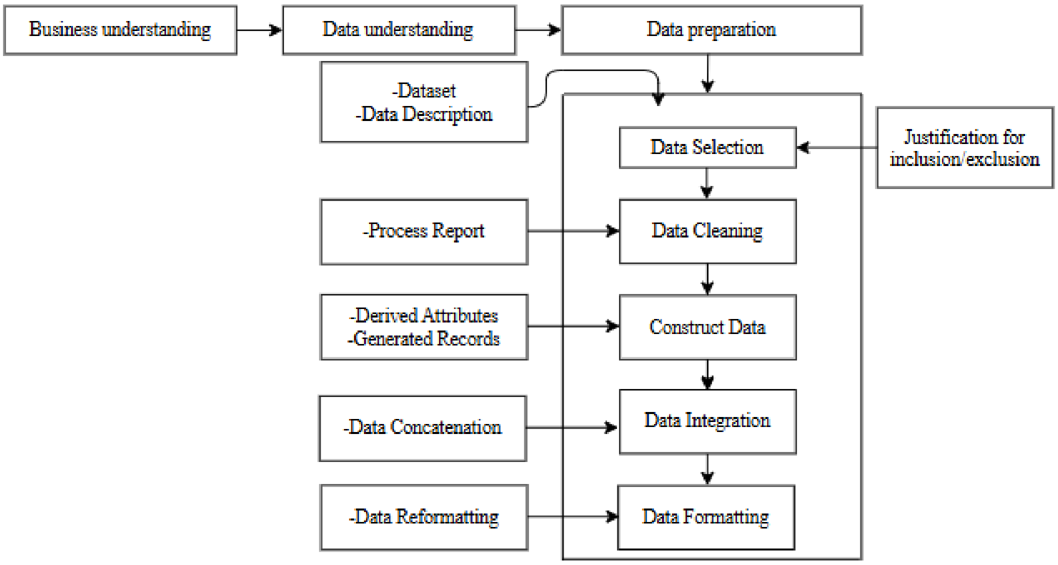
The data preparation phase (Fig. 3) is one of the most resource-intensive stages in the standardized data extraction method. This phase emphasizes the necessity for adequate time management.

The provided data sets are organized and described in a manner proper to the purposes of the process. Both the strengths and weaknesses of the characteristics of the available information set are identified. These steps serve as a transitional stage between understanding the data and preparing it for subsequent use.

The analyzed datasets provide the process manager with essential information for making informed decisions about which data should be included in the overall dataset for further analysis. At this stage, additional justification is required to determine whether to include or exclude specific datasets from the main set of information.

After preparing the final data set, the next step is to clean it by removing any information that is not suitable for modeling. It is possible that some of the collected data has been synthesized or intended for different purposes. This requires specifying and formatting the data so that it can be effectively used in the subsequent steps of modeling.

Figure 3 Structure of the data preparation process



Source - own

In the data preparation phase, alongside the previously discussed stages of data selection and cleaning, it is important to identify the need to generate additional features and characteristics of the data. Key considerations include the proper preparation of the dataset, its analysis, and recognizing the necessity to create new features derived from existing ones. This process involves the combination, synthesis, and selection of specific characteristics.

It is also essential to determine whether additional records must be generated from the existing data. This can be achieved through a random selection of available rows and their duplication.

After completing these tasks, the data integration process begins. This involves merging different datasets to create a single unified information array. If necessary, further data reformatting is performed to ensure the information is in a more convenient and suitable form for modeling.

Modeling

The modeling process (see Fig. 4) is the fourth key step in the standardized industrial process for extracting knowledge from data. Essentially, modeling involves a sequence of steps and techniques that use algorithms to develop a system for knowledge extraction from data. In this method, the modeling process consists of several main steps: selecting a modeling technique, designing a test plan, constructing the model, and evaluating its performance.

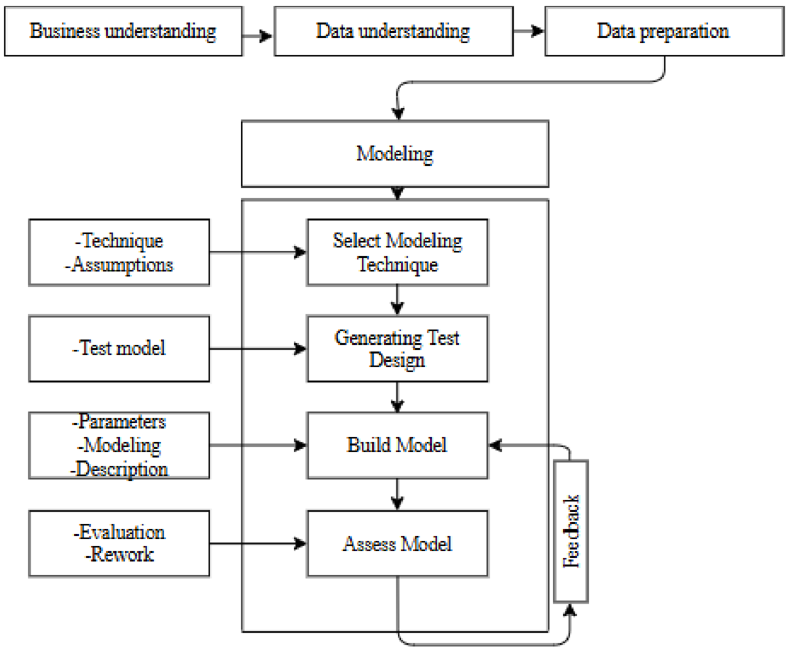
Selecting a modeling technique involves identifying the appropriate methods based on the organization's needs, the data available, and the competencies of the modeling team. A list of appropriate techniques is created, identifying the strengths and weaknesses of the tools under consideration. Assumptions are made regarding the methodology being studied, the expected outcomes, and the need to enhance the existing tools. A modeling method that meets the intended objectives is selected.

Once an informed choice about the modeling method is made, the next step is to build a test model. Constructing a test environment for modeling is essential for validating the suitability of the chosen techniques. This phase focuses on identifying potential weaknesses in the selected methods and exploring ways to improve the model's performance.

The testing phase and the identification of potential inaccuracies establish the prerequisites for progressing to the actual modeling phase, which is model building. The model development occurs in several key stages.

- The first stage is related to the determination and adjustment of the parameters. The accuracy of the model's outcomes is dependent on this phase.
- The second stage focuses on the actual modeling process, which includes the intentional selection, combination, and integration of various algorithms.
- The third stage is dedicated to describing the developed model. This includes preparing detailed documentation for each component within the model, outlining the processes involved and the algorithms utilized.

Figure 4 Structure of the modeling process



Source - own

After completing the modeling phase, the next step is to proceed to verify the results and assess the quality of the model. If the results meet the established criteria - advance to the next phase. If they do not - need to identify the gaps in the modeling process by reviewing all the steps taken, from setting the coefficients to the final phase.

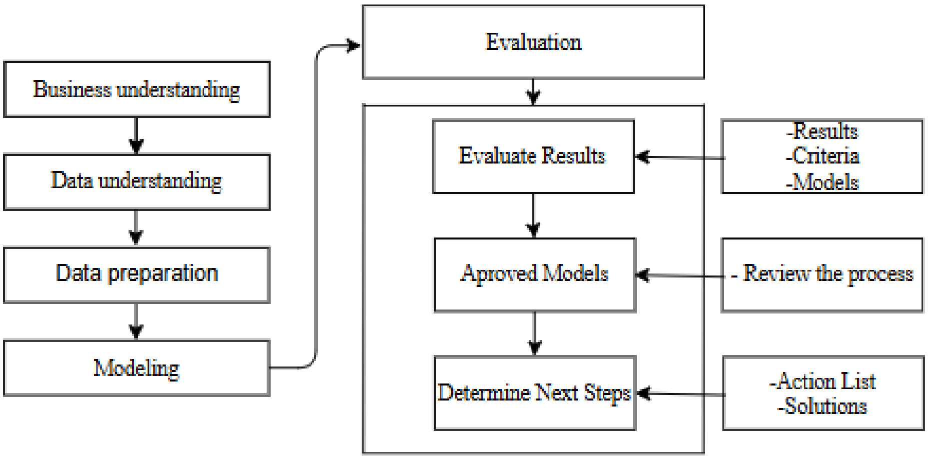
Evaluation

The evaluation stage in the standardized method for extracting knowledge from data involves a thorough assessment of the previous phases. To ensure accurate analysis and evaluation, it is essential to determine whether the results obtained during the first four stages are meaningful and suitable for business objectives.

The evaluation process (Fig. 5) consists of three main phases: assessing the results, reviewing the process, and determining the next steps. During the results evaluation phase, a thorough analysis of the outputs generated by the model is conducted. The relevance and adequacy of these results are assessed based on the established goals and the specific business area. The statistical significance of the results is determined, along with their relevance to the overall process. In the second step of the evaluation, the relevance of the established business success criteria and the adequacy of the measurements used are monitored. The results obtained from the validated models are then evaluated and compared.

The next phase of the process involves a detailed breakdown and analysis of the workflow. Monitoring is carried out on all steps and components of the model to identify any potential weaknesses in the planned sequence of work.

Figure 5 Structure of the assessment process



Source - own

The third phase of the process focuses on determining the next steps. Based on the results from the previous two phases, a detailed list of potential actions is created. If the outcomes from these stages are unsatisfactory, the list includes guidelines for improving the actions at any necessary phase. Conversely, if the results meet the initially established criteria and standards, then they are implemented accordingly.

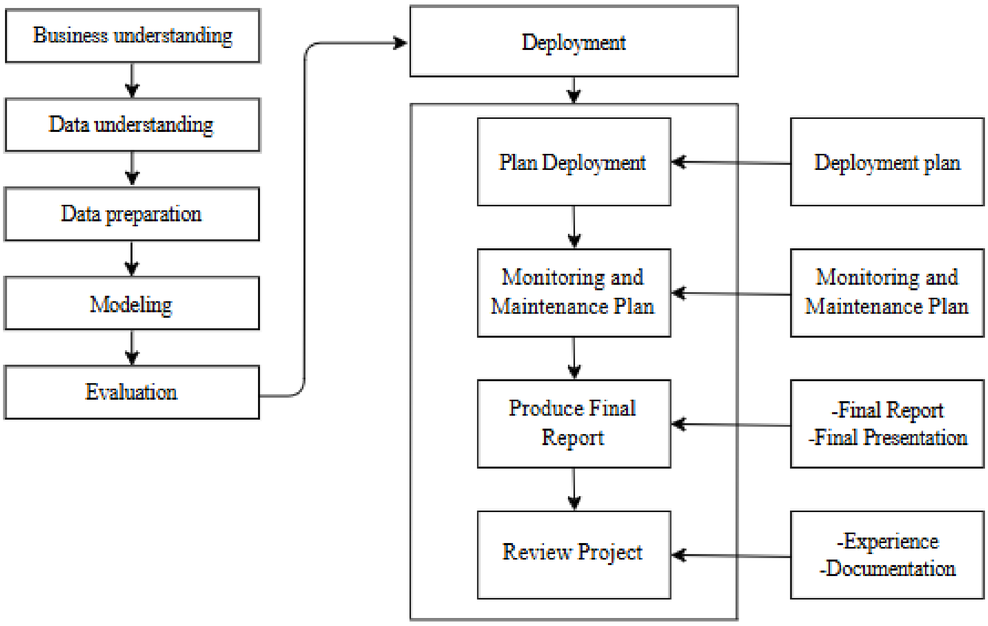
Deployment

The sixth and final phase of the CRISP-DM method focuses on implementing the new knowledge gained and integrating it into the organization's processes. This phase consists of four key stages: creating a deployment plan, developing a monitoring and maintenance plan, preparing a final report, and conducting a project review (see Fig. 6)

When preparing a plan for the deployment and implementation of the knowledge acquisition process, it is crucial to consider the need to improve and optimize work processes within the organization. This ensures the effective integration of newly developed processes.

To successfully integrate these new work processes, it is essential to establish fundamental criteria for monitoring and maintaining the system. Additionally, plans should be created for regular review of the system. The frequency of monitoring and maintenance will depend on the complexity of the model developed, the intricacies of the data extraction process, and the specific requirements of the business, among other factors.

Figure 6 Structure of the implementation process



Source - own

After establishing the plans for implementing and maintaining the process, a final report is prepared to summarize the work completed. This report outlines the steps in the process along with their descriptions. Additionally, a suitable visualization of the actions taken is created.

The final phase of implementing the standardized method for extracting knowledge from data, known as CRISP-DM, involves reflecting on the experience gained throughout the project. This includes compiling detailed documentation for each part of the process.

Individualization through a clustering algorithm (CRISP-DM implementation)

Business understanding

When creating individualized financial instruments, it is essential to adopt a comprehensive approach that maximizes the level of customization possible. However, economic feasibility does not necessarily mean designing unique products for every individual. A feasible method for addressing this challenge is clustering. This process involves segmenting individuals into clusters based on their similar characteristics, using a clustering algorithm. By grouping individuals into larger categories, we can achieve personalization at the group level.

This method identifies several key groups of variables that facilitate individualization through the analysis of large data sets. These key groups are established based on prior knowledge during the model construction process. The study is centered around the following main groups:

Demographic Variables. This group includes key demographic data relevant to building an individualization model. The aim is to establish the relationships between these indicators and to differentiate among individual groups.

Socio-Economic Variables. This refers to constructing a comprehensive profile of individuals to aid in the distinction of different groups. It focuses on financially active participants and their characteristics.

Psychological and Personal Variables. A crucial aspect of building an individualization model involves variables related to psychological traits and personal characteristics.

Financial and Banking Variables. This group emphasizes the need to construct a profile of active users of financial services. It analyzes individual behavior regarding the active use of financial products.

Personal Preferences. A foundational element in developing individualized models, this category encompasses both clients' desires and the challenges and uncertainties they encounter.

Individual Risk Preferences. This is a strictly personal measure reflecting how individuals respond to uncertainty in investments. This indicator is dynamic and varies according to individual characteristics and personal experiences.

One of the key elements in creating personalized investment instruments is understanding an individual's personal preferences. Along with information about specific investment options, it is essential to gather data on the personal and financial goals of the individual to achieve a more effective and tailored investment experience. This comprehensive approach combines various characteristics, encompassing both the individual's preferences for investment instruments and insights related to their financial objectives.

Data understanding

In the context of big data in financial services, there are fundamental challenges related to the complexity of the data and legal regulations, such as the Personal Data Protection Act, the Bank Secrecy Act, the GDPR, and others. The methodology for acquiring data to support the model involves generating information arrays.

Collecting data through standard survey methods is a complex and time-consuming process. The specificity of the required information often leads to negative responses from respondents.

Given these challenges, an alternative approach is needed to gather the necessary data. One effective method is to simulate a synthetic database that includes a variety of demographic, personal, individual, and banking variables and indicators.

Due to the specificity requirements of the empirical study, it is essential to generate this database using a combination of simulation techniques. The individual data components are combined through a process known as probabilistic concatenation

A variety of software and algorithms based on different data simulation methods are employed. The data simulation process consists of several key phases:

- Variable selection
- Choice of the simulation approach
- Data simulation
- Analysis
- Validation

This topic is thoroughly explored by the authors of this study (Marchev, Marchev, 2023a, 2023b; Marchev et al, 2023).

Data preparation

The variable transformation phase involves modifying a portion of the initially collected data to create new variables or to make existing ones more recognizable by models that use this information.

For the current information array, a transformation of the type is required from absolute values into categories that correspond to specific ranges. The first step in preparing the data for analysis is to code each column in a manner suitable for the software being used.

Table 1 Sample of variable encoding performed

Variable	System code
Sex/Gender	sex
Age - completed years	age
Level of education	lv_educ
Employment status	empl_stat
Marital status	marit_stat

Source: own

Additionally, the available data requires further processing. Along with processing the available information, it's essential to convert all non-numerical data into corresponding numerical indicators. During the data preparation phase, steps should be taken to maintain the statistical distributions derived from the original information sources. If needed, groups of variables with a small number of records can be combined with others.

Modeling

Individualization as a product is achieved through a clustering process applied to the database created in the previous step. The result of clustering is a collection of groups, or clusters, which are distinguished based on the variables used and the similarities among the individual records.

A fundamental aspect of the individualization process through clustering is the quality of the available information. To ensure this quality, data that has been previously prepared, cleaned, and transformed is utilized. Necessary scalar measures are determined for each factor included in the upcoming process.

The information set is divided into a training set and a test set by specifying the proportions for each part. For this research, the dataset is divided as follows: 80% of the data is allocated to the training set, while the remaining 20% is designated as the test set.

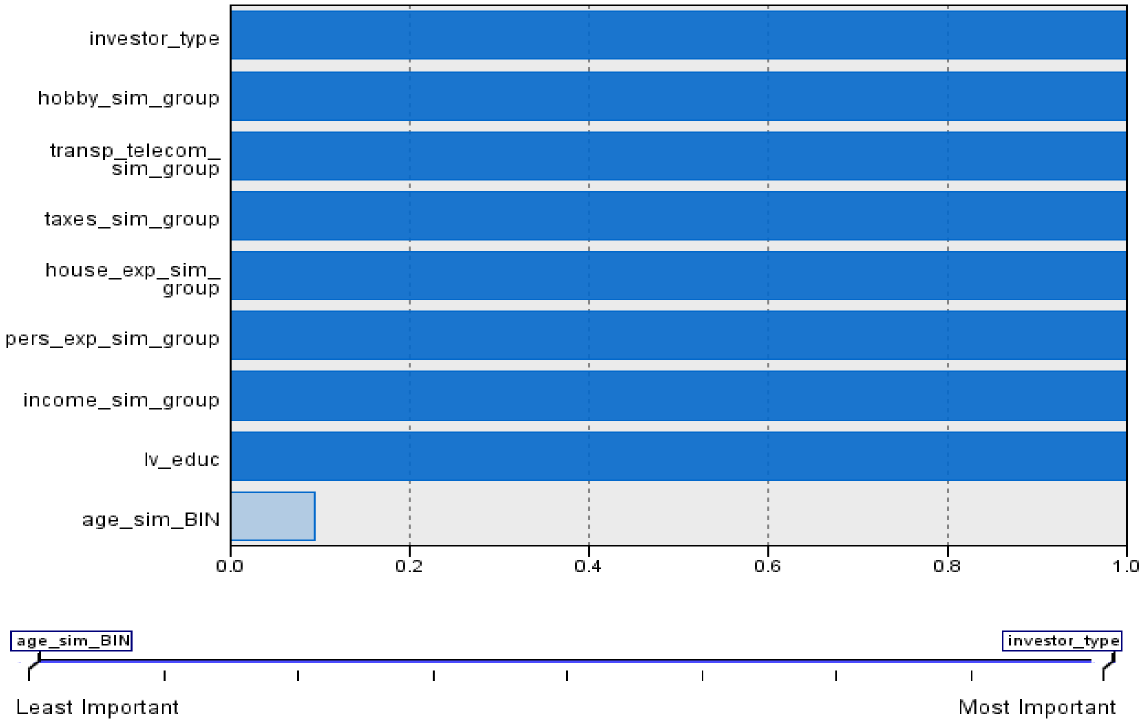
Building a clustering model involves a series of sequential steps. After determining the training and test subsets, the selection of a working method is carried out after testing various approaches to identify the most effective one. Multiple methods have been observed in the clustering process. Marchev (2021) provides more detailed information on the matter. This selection process considers the specific characteristics of the data, and the goals set for the research.

Evaluation

The analysis of the individualization model through cluster generation involves identifying and monitoring the characteristics which are with most statistical significance, based on the selected method.

The data most influential in this process can be found among individual characteristics, demographic features, and personal qualities of individuals.

Figure 7 Analysis of the characteristics



Source: Own

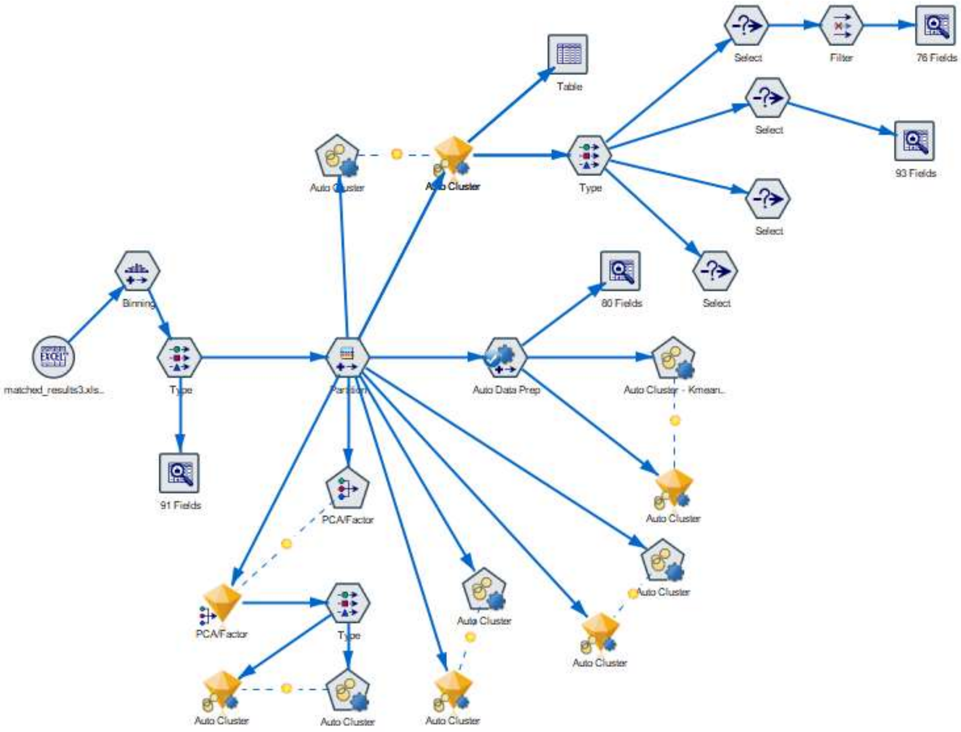
The quality of the chosen method for conducting the study is crucial. It is important to note that the effectiveness of the clusters is directly related to the appropriateness of the selected method for generating the necessary large data sets. If the chosen tool for identifying individual groups is not suitable, it may result in the creation of many individual clusters that exhibit relatively lower quality in the clustering process.

Deployment

The implementation of an individualized model through clustering can be carried out using various software tools. The most used options for this purpose include SPSS Modeler, MATLAB, Python, and R/R Studio.

Each of these tools has its own advantages and disadvantages, depending on the specific needs of the model being developed. When selecting software, it is important to consider both the expertise of the research team, and the time required to complete the work.

Fig. 8 Cluster model diagram



Source: own

CONCLUSION

In conclusion, integrating the Cross-Industry Standard Process for Data Mining (CRISP-DM) into the FinTech sector marks a significant advancement in leveraging big data to enhance user experiences and optimize financial services. This approach provides a framework to navigate the dynamic environment of the FinTech industry, enabling a more comprehensive understanding for customers, regulators, and potential sector investors. The structured methodology of CRISP-DM facilitates a thorough grasp of business objectives, data characteristics, and modeling processes essential for effective knowledge extraction.

The evolving behavior of individual investors, combined with technological advancements, has created a favorable environment for FinTech companies to thrive by offering personalized financial solutions tailored to specific consumer needs. By systematically implementing this methodology, FinTech companies can improve service delivery and foster greater customer satisfaction through customized investment strategies. The dynamic nature of the financial system provides a solid foundation for the adoption of such data-driven approaches. Furthermore, these actions are becoming essential for maintaining competitiveness and meeting the ever-changing demands of consumers in the financial services market.

References

Adriaans, P.; Zantinge, D. (1996). Data Mining. Addison-Wesley. Harlow, England

Allport GW. (1961). “Pattern of growth in personality”, NY: Holt, Rinehart and Winston

Brachman, R.J.; Anand, T. (1996). The Process of Knowledge Discovery in Databases: A Human-Centered Approach. In Fayyad, U.M., Piatetsky-Shapiro, G., Smyth, P., & Uthurasamy, R. (eds.). Advances in Knowledge Discovery and Data Mining. AAAI /MIT Press

Brown, M., What IT Needs to Know About the Data Mining Process, Forbes, Jul 2015

Chaudhry, S., Ahmed, R., Huynh, T., Benjasak, Ch., 2022, Tail risk and systemic risk of finance and technology (FinTech) firms, Technological Forecasting and Social Change Volume 174, January 2022, 121191, https://doi.org/10.1016/j.techfore.2021.121191

Cheng, A., 2023, Evaluating Fintech industry's risks: A preliminary analysis based on CRISP-DM framework, Finance Research Letters Volume 55, Part B, July 2023, 103966, https://doi.org/10.1016/j.frl.2023.103966

Costa Jr PT., McCrae R., R., Dembroski T, 1989, "Agreeableness vs. Antagonism: Explication of a potential risk factor for CHD. In A. Siegman & T. M. Dembroski (Eds.), In search of coronary-prone behavior. Beyond Type A Hillsdale. NJ: Lawrence Erlbaum,1989; 41-63

Leaper, N., 2009, A Visual Guide To CRISP-DM Methodology

Marchev, A., Jr, Marchev, V., 2023, Automated Algorithm for Multi-variate Data Synthesis with Cholesky Decomposition, ICACS '23: Proceedings of the 7th International Conference on Algorithms, Computing and Systems, https://doi.org/10.1145/3631908.3631909

Marchev, A., Jr, Marchev, V., 2023, SYNTHESIZING MULTI-DIMENSIONAL PERSONAL DATA SETS. AIP CONFERENCE PROCEEDINGS; 2505 (1): 020012, DOI: 10.1063/5.0100615

Marchev, V, 2021, An Empirical Exploration of Data Segmentation and Clustering Methods: Insights from SPSS Modeler. Vanguard Scientific Instruments in Management, vol. 17. no. 1, 2021, ISSN 1314-0582

Marchev, V., Marchev, A., 2023,Individualized Fin-Tech Investment Services, Journal of Global Strategic Management, V. 17, N.2, 2023, December, isma.info, 025-034, DOI: 10.20460/JGSM.2024.330

Marchev, V., Marchev, A., Piryankova, M., Masarliev, D., Mitkov, V., (2023), Synthesizing an anonymized multidimensional dataset featuring financial, economic, demographic, and personal traits data, VSIM, vol. 19, no. 1, 2023, ISSN 1314-0582

Mhlanga, D., 2024, Big Data and Financial Technology (Fintech) Towards Financial Inclusion http://dx.doi.org/10.2139/ssrn.4697365

Reinartz, T., & Wirth, R., 1995, The Need for a Task Model for Knowledge Discovery in Databases. In: Kodratoff, Y., Nakhaeizadeh, G., & Taylor, C. (eds.). Workshop Notes Statistics, Machine Learning, and Knowledge Discovery in Databases. MLNet Familiarization Workshop, April 28-29, Heraklion, Crete, pp. 19-24

Schröer, C., Kruse, F., Gómez, J., 2021, A Systematic Literature Review on Applying CRISP-DM Process Model, CENTERIS - International Conference on Enterprise Information Systems / Pojman - International Conference on Project Management / Hcist - International Conference on Health and Social Care Information Systems and Technologies 2020, Procedia Computer Science 181 (2021) 526–534

Shearer, C., 2000, The CRISP-DM Model: The New Blueprint for Data Mining, Journal of Data Warehousing,

Wirth, R., Hipp, J., 2000, CRISP-DM: Towards a Standard Process Model for Data Mining, Proceedings of the 4th international conference on the practical applications of knowledge discovery and datamining (4), pp 29-39