TECHNICAL SCIENCES

IMPROVING THE DECISION-MAKING SYSTEM BY APPLYING MACHINE LEARNING ALGORITHMS IN A TELECOMMUNICATIONS COMPANY

Ryzhov A.
Georgia, Adjara region, city Makhinjauri, str. Israfil Jincharadze, h.3, app. 611
https://doi.org/10.5281/zenodo.10629232

Abstract

Currently, artificial intelligence (AI) is actively used in many fields, including the telecommunications industry. Machine learning (ML) algorithms play a key role in automation, classification, segmentation, and clustering tasks to support decision-making systems (Jepsen 2021). ML algorithms are introduced to achieve classified product offers in the accounting system, thereby building the company's financial statements. This study solves the problem of automating a business process by using ML algorithms.

One of the most significant technological advances of the last decade has been the development and widespread adoption of ML. Decision-making systems using ML algorithms can be applied in various areas for intelligent decision-making to minimise manual human labour in business processes.

Various tasks were identified to achieve the study's goal. They included analysing the company's existing business processes, identifying bottlenecks and shortcomings, developing KPIs to measure its effectiveness, and proposing a new solution for organising the business processes.

This research proposes a model that automates manual labour. This model is a technical solution for making changes to an existing information system, supporting the operation of an ML model, and using the prediction results in an automated business process.

Keywords: machine learning, decision-support system, improving information system, ML models, classification, ML algorithms, refactoring business process, artificial intelligence

Introduction

This study considers the possibility of introducing ML algorithms and making changes to the existing information system to partially automate the business process in a telecommunications company.

In the company under study, classifying services (product offers) is a regular process since the frequency of new services is about 100 units per month. The company has a certain list of service classes. This list is not static; new classes may appear in it as decisions are made to create new products. Each new product offering must be assigned to one of the existing classes. This information is used to build official financial statements to correctly allocate income from services to income statement items.

When deciding whether a new service belongs to a particular class, a list of attributes and characteristics of product offerings is considered (service types include voice services, data transfer services, and roaming). To perform a resource-intensive process of classifying services, knowledge of the subject area is required. Currently, experts perform this job.

This research proposes automating the classification stage using ML algorithms. The introduction of ML algorithms into the contour of an existing information system implies a change in the logic of business processes and the information system at the data storage organisation level. It also implies changes in organising interaction with built models and data flow between elements of the information system.

Thus, this research aims to discover how to improve the classification of product offers, which is resource-intensive. To achieve this goal, the following tasks were performed:

- Analysis of the existing business process, identifying bottlenecks and proposing changes.
- Analysis of the existing information system and proposing changes at different levels.
- For automation, building a classification model based on ML algorithms.
- Developing reference terms to outline the organisation of work with the ML model, some use cases of implemented interfaces.

The object of the research is the company's information systems involved in the business process of classifying product offerings. The subject of the research is the automation of the business process of classifying product offerings.

This study's hypothesis is that the introduction of an ML algorithm will increase processing speed and reduce costs without loss of accuracy.

This research consists of an introduction, three additional sections, a conclusion, references, and a list of tables. The first additional section contains an overview of the experiences using ML to automate company business processes, a description of ML algorithms, and the algorithm choice results we used to conduct modelling experiments. The next section contains a general description of the business processes of companies in the telecommunications industry and the current and modified researched business process for classifying product offerings. The last section describes the procedure for developing and implementing an ML model into companies' information systems. It also includes a description of the data used to train the model, experiments and results of model training, and necessary actions to implement the model in the company's information system under study.

Despite the fact that the study was conducted to solve the problem of automating the classification process in a particular telecommunications company, the results of this study may be interesting for use in different organizations when solving the problem of automating classification.

Experience in implementing ML models in automating production business processes

Experience of automating business processes

Currently, ML is actively used in corporate software to automate routine business processes. At the same time, the degree of improvement in business processes results using ML can be significant, depending on the application area. ML algorithms are used to identify hidden patterns in data, build predictive models, optimise business processes, and automate routine operations and decision-making in complex areas. ML algorithms are also used to automate work with documents and multimedia data (AI Trends 2021).

Considering the experience of using ML in the telecommunications industry, researchers must consider that these companies collect vast amounts of network and customer data. The data's volume and complexity are rapidly increasing due to the growing number of devices and the extent of customer data. ML is becoming necessary to analyse and use this data. Telecommunications operators use ML to help network planners and engineers effectively manage alarms and perform other routine operations.

Since there are many non-automated business processes (BP) in telecommunications companies, the following reviews their experience using ML.

Consider the BP recognition of checks (receipts) and intelligent invoice processing: The BP ML algorithm checks the scanned receipts and determines the type of receipt, then automatically matches it with a ledger entry. During intelligent invoice processing, the ML algorithm analyses the invoice number and items and compares them with the corresponding purchase order. As previously stated, when considering the telecommunications industry's experiences with ML use, it is necessary to address the fact that today, these companies collect increasingly huge amounts of complex network and customer data. Efficient use and analysis of this data makes ML even more necessary.

Telecommunications operators use ML to help network planners and engineers effectively manage alarms and perform other routine operations. For example, in the United States, AT&T uses ML to create a virtual world that describes the external environment (poles, buildings, and other objects) to determine where base stations can be placed without having to visit a specific facility. Telecom operators in China are already using artificial intelligence (AI) as they move towards autonomous networks.

In October 2019, GSMA published a report (GSMA, 2019) on several cases of AI use in the Chinese market: the creation of an intelligent robot planner, analysing the root causes of wireless alarms, intelligent management of transport network slices, and intelligent identification of services.

China Mobile Experience

Alarm Aggregation – China Mobile received 600,000 packet transmission alarms daily, so the company started looking for new ways to troubleshoot its mobile network. The company has implemented an incident-oriented approach to alarm management that flags the root cause of alarms, rather than all related warnings, with an ML-based aggregation solution. The result was a compression of alarms to 99% (McElligott 2020).

China Unicom Experience

VoLTE Packet Loss (Voice over LTE) – China Unicom has applied AI-based optimisation of mobile radio parameters to reduce the VoLTE packet loss rate by 5%. In this case, an ML algorithm was used to establish a correlation between the user's experience with the VoLTE service and network performance indicators. This allowed China Unicom to optimally adjust the radio communication parameters. The company also reduced the time required for troubleshooting to one minute (McElligott 2020).

Turk Telekom Experience

Turk Telekom is a telecommunications operator providing fixed, broadband, and mobile communication services in Turkey. Tombes (2020) discussed the widespread use of ML in the company's business processes. Turk Telekom has deployed ML on a big data platform to identify anomalies and reduce potential problems affecting customers in the network's backbone. This deployment used the following algorithms: K-nearest neighbour (KNN), Gaussian Mixture Model (GMM) for a probabilistic approach, statistical methods (e.g., Z-estimation), and Isolation Forest for uncontrolled detection of outliers.

Turk Telekom has implemented anomaly detection scenarios divided into three main categories: receive-transmit time (RTT), network traffic, and alarm duration. All three provide improved visualisation of network performance and opportunities for preventive network maintenance.

RTT (English round-trip time) delays. Starting in April 2019, Turk Telekom deployed six RTT delay anomaly scenarios using three measurement applications (Sigos, Gezgin, and TWAMP). The leading division of Turk Telekom analysed RTT data about 1 GB in size over the next eight months. It shared 1,105 detected anomalies with network operations, which made corresponding improvements. As a result, Turk Telekom reduced RTT delay values by 8%, 5%, and 9%, respectively, in peer-to-peer, international, and mobile channels. This RTT latency reduction ultimately benefits customers. Receiving warnings even about minor anomalies using ML-based analytics allows network teams to carry out preventive maintenance before RTT delays reach critical thresholds and the quality of service deteriorates.

Network traffic. Turk Telekom has also deployed three anomaly detection scenarios focused on network traffic values based on devices, services, and applications. Starting in May 2019, the lead department analysed approximately 20 GB of network traffic data and found 1,800 anomalies, which it shared with the relevant operational teams to improve customer service quality.

Specific anomalies determine the reaction. On the one hand, when application traffic drops, Net Ops first checks to see if the application server has shut down; then, it checks to see if something has gone wrong on the network. On the other hand, a significant increase in network traffic often indicates a malicious attack. They can also point to customers exceeding the agreed limits.

Telekom Malaysia Berhad Experience

This Malaysian telecommunications company has developed and implemented a network diagnostics and expert advice system for the IDEAS+ support service to help operators quickly solve customer problems. The operation of this system is based on ML algorithms.

Kridel (2020) discussed ML use in the company's "IDEAS+" system. ML in IDEAS+ uses two algorithms. One of them is k-nearest neighbours, which uses existing known variables to understand a new unknown variable. The second algorithm is a naive Bayesian method that creates a hypothesis or assumption about a variable based on the available data. For example, knowing that most people do not like cold rain, a naive Bayes could look at the current weather and assume that most would stay home. These algorithms help IDEAS+ quickly analyse a large set of variables that Telekom Malaysia operators encounter every day when processing customer requests. There are more than 500 classifications in IDEAS+. They combine data, metadata, and multimedia from various Telekom Malaysia data sources from network equipment, network performance and alarm systems, clients, billing, and client applications. Operators are now provided with the information they need to solve the problem faster than if they did the bulk of troubleshooting analysis independently. This optimisation increases customer satisfaction by reducing the time spent waiting and talking to the operator and operator efficiency in choosing the right solution (e.g., sending a specialist trained for this particular problem). The results have been quarterly savings of about $190,000, an increase in the average processing time up to three times, and an increase in troubleshooting accuracy up to 98%.

Machine learning: Choosing an ML Algorithm to Build a Model

This section defines machine learning and considers various machine learning algorithms. Machine learning is based on a set of algorithms that use the provided data to train and predict, optimise a utility function under uncertainty, recognise hidden structures in the data, and classify the data with a brief description. As new data comes in, these algorithms learn and optimise their performance, improving performance and developing intelligence over time. The following algorithms are considered: k-nearest neighbour algorithm, regression models, clustering, dimensionality reduction algorithm, reinforcement learning, Random Forest, Decision Tree, and XG-Boost. Supervised ML is used to solve classification problems. Thus, the list of available methods to work with and conduct experiments on is narrowed. This study used the following ML algorithms to classify the product offerings of a telecommunications company with a large number of attributes (38):

- Decision Tree
- Random Forest
- XGBoost

To successfully implement the created ML model into the contour of the existing information system, this research analyses the existing business process of launching and classifying product offers and proposes changes in its implementation.

Description of business processes and suggestions for automation

This section considers the business processes of the telecommunications industry as people manually perform a large number of repetitive and rules-based actions.

General description of business processes in the telecommunications industry

A detailed consideration of the business process ‘Development, launch, closing of a product/service’ is critical since it contains the following sub-processes involving manual classification of product offerings:

- Development of subscriber equipment.
- Development, launch, and implementation of marketing programs.

The sub-process ‘Development, launch, implementation of marketing programs’ consists of several sub-processes, but the sub-process ‘Product development’ is the most interesting to consider since it carries out product classification.

The AS IS diagram developed in BPMN notation presents the complete business process of classifying services, as seen in Figure 1. Here, the main operations are highlighted, and information flows, roles, divisions, and the company's information system elements used are identified.

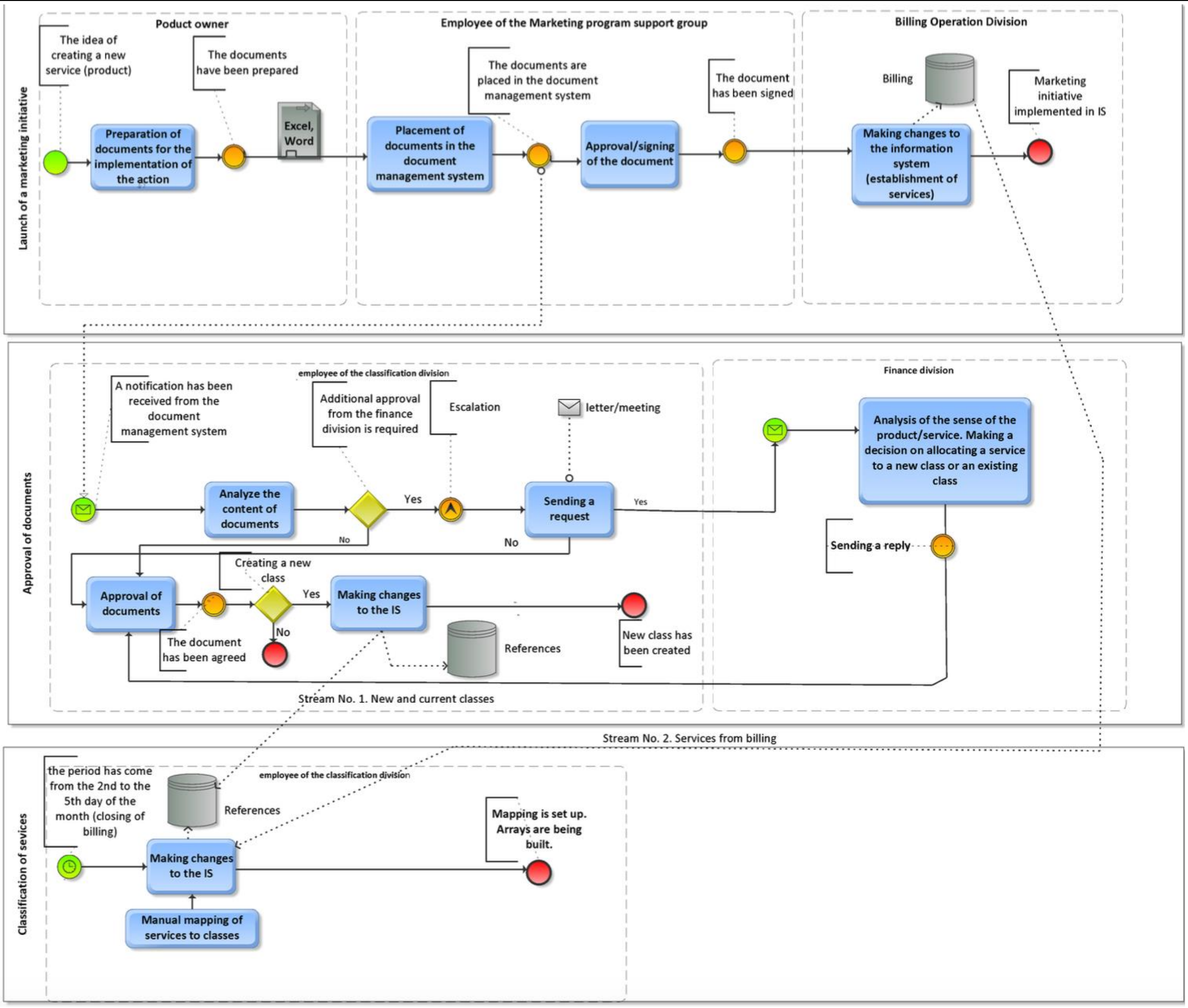

Figure 1. Diagram of the Product Offering Classification (AS IS) process.

As a result of the study, the authors propose reengineering the current business process of the telecommunications company studied. Figure 2 shows a modified business process model (TO BE). The blue colour indicates new elements of the business process and changes in the information system.

When making changes to the business process, it is impossible to bypass the existing information system since the proposed changes also affect it.

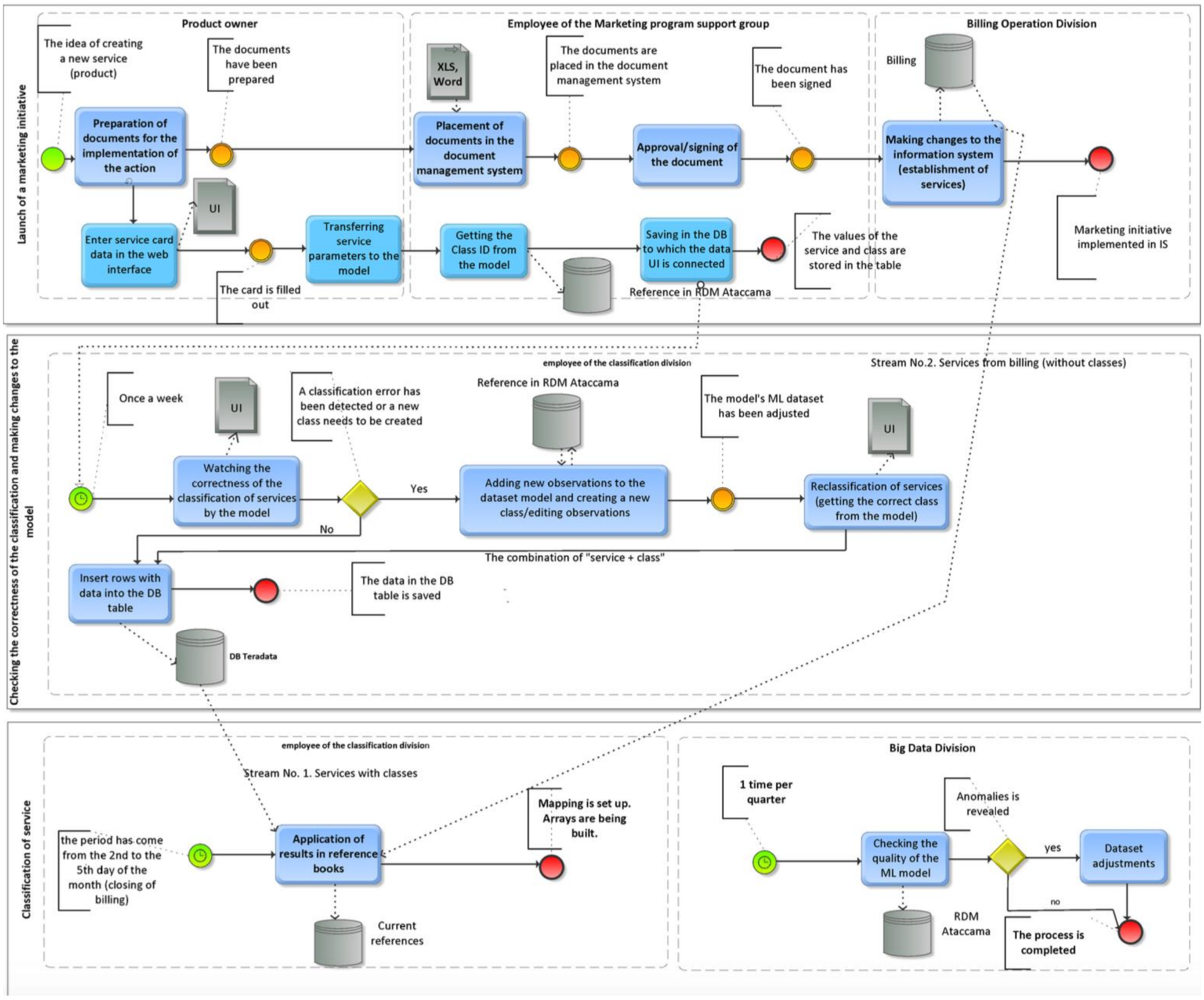

Figure 2. Product Offer Classification (TO BE) Process Diagram.

Solution of the problem of business process automation using ML algorithms

This section confirms the study's hypothesis that using ML algorithms will allow automation of the selected business process. Testing this hypothesis required going through the complete ML cycle, discussed below.

Description of the building procedure for an ML model and its implementation results

Business analysis was performed first. At this stage, the main business variables of the model used for the forecast were identified. This involved forecasting service classes (chart of accounts lines for the company's financial reporting system). At this stage, we also decided to use business metrics to evaluate the model's implementation and to forecast error risks and the required metrics for assessing the quality of the model (recall, F1, precision, and other metrics).

The next step was data analysis and preparation. This stage includes several tasks: data analysis and collection, data normalisation and modelling, and feature construction (class parameters).

The next step of the ML model-building procedure was modelling, one of the main stages. This stage can be represented by several tasks: choosing ML algorithms for the model, training the ML model, and evaluating the results. At this stage, the data set's distribution percentage for the test and training samples is determined — 70% for the training and 30% for the test. In this case, the training set was used to train the model, and the test set was needed to obtain the model quality values. The models were trained iteratively, and the prediction result was recorded after each iteration. Further, the results of the work of each ML algorithm were summed up.

The final task was evaluating the results, which involved analysing and selecting the best model.

Description of the data used to train the model

This section discusses the classification process purpose and the available data characteristics.

The classification of services by lines in the chart of accounts is necessary for constructing the company's official financial statements, analysing income from services, and distributing income among different RAS (Russian Accounting System) accounts. Based on the data content analysis in the database management system tables, a data set comprising 38 variables and 540 lines with the group_id field identifying the class. There are 10 observations per class. Unique classes – 55.

The class parameter data in the dataset was created in a binary type. Figure 3 shows the distribution of 1 and 0 by each variable.

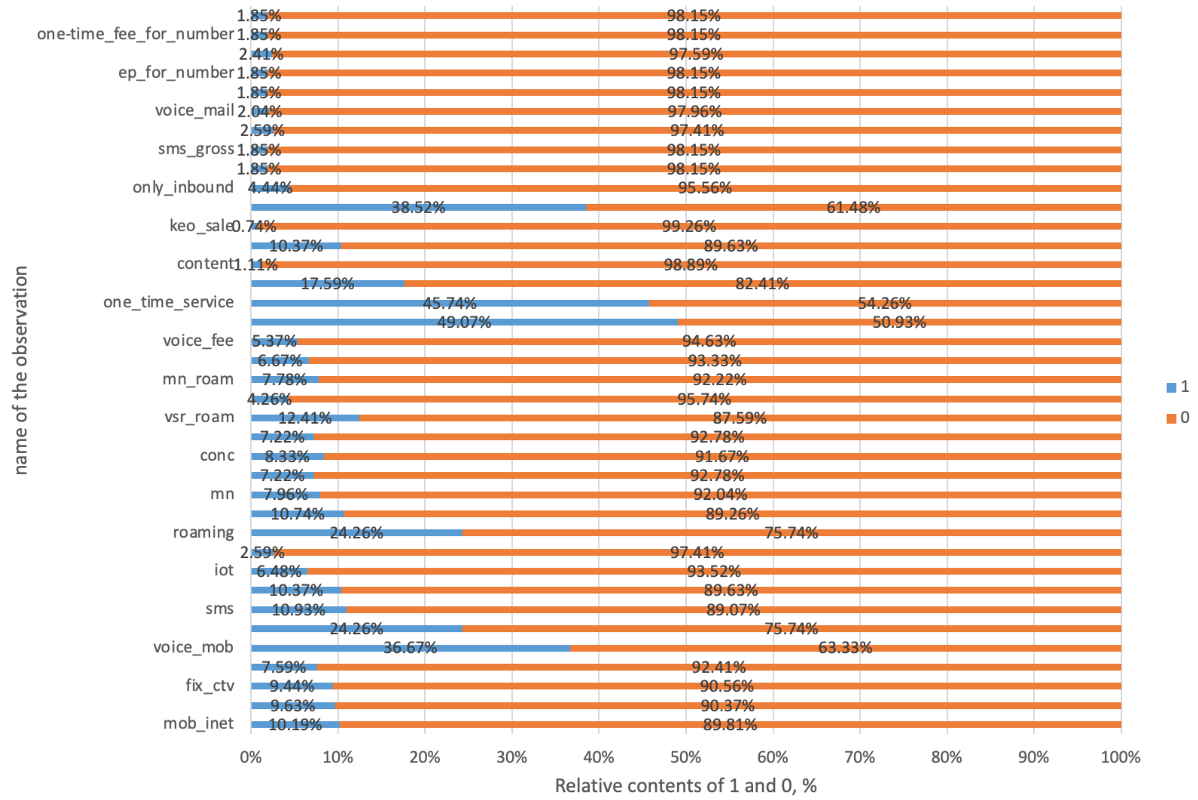

Figure 3. Distribution of values 1 and 0 in the data set, by variables.

Description of experiments

As part of the research, three experiments were conducted to build models using three algorithms: Decision Tree, Random Forest, and XGBoost.

The score_value metric was used to evaluate the performance of the models. score_value is the average accuracy for the given test data.

$$score_{value} = \frac{\sum \quad accuracy\ n\ classes}{n\ classes}, \qquad (1)$$

where n is the number of all classes in the training sample.

The average accuracy (function accuracy_score) calculates the number of correct predictions. The minimum value is 0 and the maximum is 1 if the entire set of predicted labels for the sample fully corresponds to the true label set. The average accuracy is calculated according to the following formula:

$$accuracy = \frac{tp}{ap}, \qquad (2)$$

where tp is the number of correct predictions and ap is the number of all predictions.

There are also used metrics for assessing the quality of the models as follows:

- precision
- recall
- f1 score

We consider these metrics and the formulas for their calculation in more detail below.

Precision is the classifier's ability not to label negative samples as positive and is calculated according to the formula:

$$Precision = tp \,/\, (tp + fp), \qquad (3)$$

where tp is the number of true predictions and fp is the number of false positive predictions.

Recall is a weighted average of each class's accuracy. This metric calculation is carried out according to the following formula:

$$Recall = tp \,/\, (tp + fn), \qquad (4)$$

where tp is the number of true positive predictions and fn is the number of false negative predictions.

The *F1-score* allows the researcher to find a balance between accuracy and recall. The calculation is carried out according to the following formula:

$$F1 = \frac{Precision * Recall}{Precision + Recall}, \qquad (5)$$

The minimum threshold for score_value was set to 0.8. Additionally, an expert assessment of the current level of errors was carried out. The error rate is 0.26 (out of 100 randomly selected services, 26 of them had an incorrect class).

Comparing training outcomes and choosing the best model

The decision to choose the best model was carried out according to the aggregated score coefficient. Table 1 shows the results of the aggregate score.

Table 1.

Score values by ML algorithms

Name of ML algorithm	Score_value of the test set
Random Forest	0.84
Decision Tree	0.79
XG-Boost	0.80

Based on the results obtained during the experiments, one can make an assumption about which algorithm will be preferred: Random Forest, having the highest score on the test sample. However, it is also necessary to consider several factors influencing the best model choice. Each algorithm used in the experiments had its advantages and disadvantages. Here are some of them:

- In the XGBoost model, the gradient is taken into account for each tree, and the calculation is faster than for Random Forest.
- In the Random Forest model, there are many trees with leaves of the same weight, thus increasing the accuracy of class predictions.
- XGBoost builds one tree at a time, and all data related to the decision tree is considered. If there is no missing data, it is filled.

Considering the above advantages and disadvantages of the considered algorithms and the importance of the least error probability in the service classifications, preference was given to the Random Forest algorithm since it showed the best score value on the test sample: 0.84 during experiments.

Necessary actions for loading an ML model into the contour of an existing information system

Implementing an ML model in a production environment implies the model's availability for other company systems. When loading a model, if other systems send data to the model and receive forecasts from it, then the company's systems consume and use these forecasts. The loaded model can be accessed through various interfaces, such as the Open API.

In addition to loading the model, some actions are required to use the model successfully in the service classification business process:

- Description of the new architecture: According to the architectural scheme in Figure 4, several key systems are distinguished.
- Existing: automated payment system (billing), accounting system, and RDM Atacama.
- New elements of the company's information system: the API interface for interacting with the ML model, creating methods API, creating new forms in the existing interface for the class editor and their parameters, server Linux.

The scheme assumes that when launching a new product offer, the product manager (an employee of the marketing department who releases new services) opens WEB UI in the browser, selects the parameters of the product offer, and clicks save. The ML loaded into the server accepts the data exposed in the form of an interface in the previous step and predicts the class.

This study proposes integrating the current element of the account information with the existing RDM Ataccama class and service directory product. It also proposes setting up an ETL process on the side of the accounting system, which will run daily and consume data about new services and classes in the catalogue. This everyday process is critical for reclassifying services, in which the editor of classes and services for a previously existing combination of 'Service and Class' applies changes (changes the class). Also, an ETL process will be launched monthly on the side of the accounting system, which will consume a class from the RDM Ataccama catalogue of classes and services for services that have just arrived from the billing system and for which the class was not previously specified in the account system catalogue 'Service – Plan line accounts'.

Within a month, an employee with the role 'Editor of parameters and classes' performs an operation to selectively check the correctness of assigning services to the class predicted by the model. To ensure this operation, the employee opens the directory of services and classes in RDM Ataccama (if the web interface exists, it is necessary to create new workspaces for this role). In the interface, the choice of services and class combinations is carried out, and the employee manually changes the class for certain combinations. Further, this change is fixed in the directory of services and classes of RDM Ataccama, a new version is opened, and the old version is closed by the current date. The combination of a service and a class that has been manually reclassified falls into a separate volatile table in RDM Ataccama. Then, the daily ETL process, scheduled on the side of the accounting system, loads the data into the reclassification table.

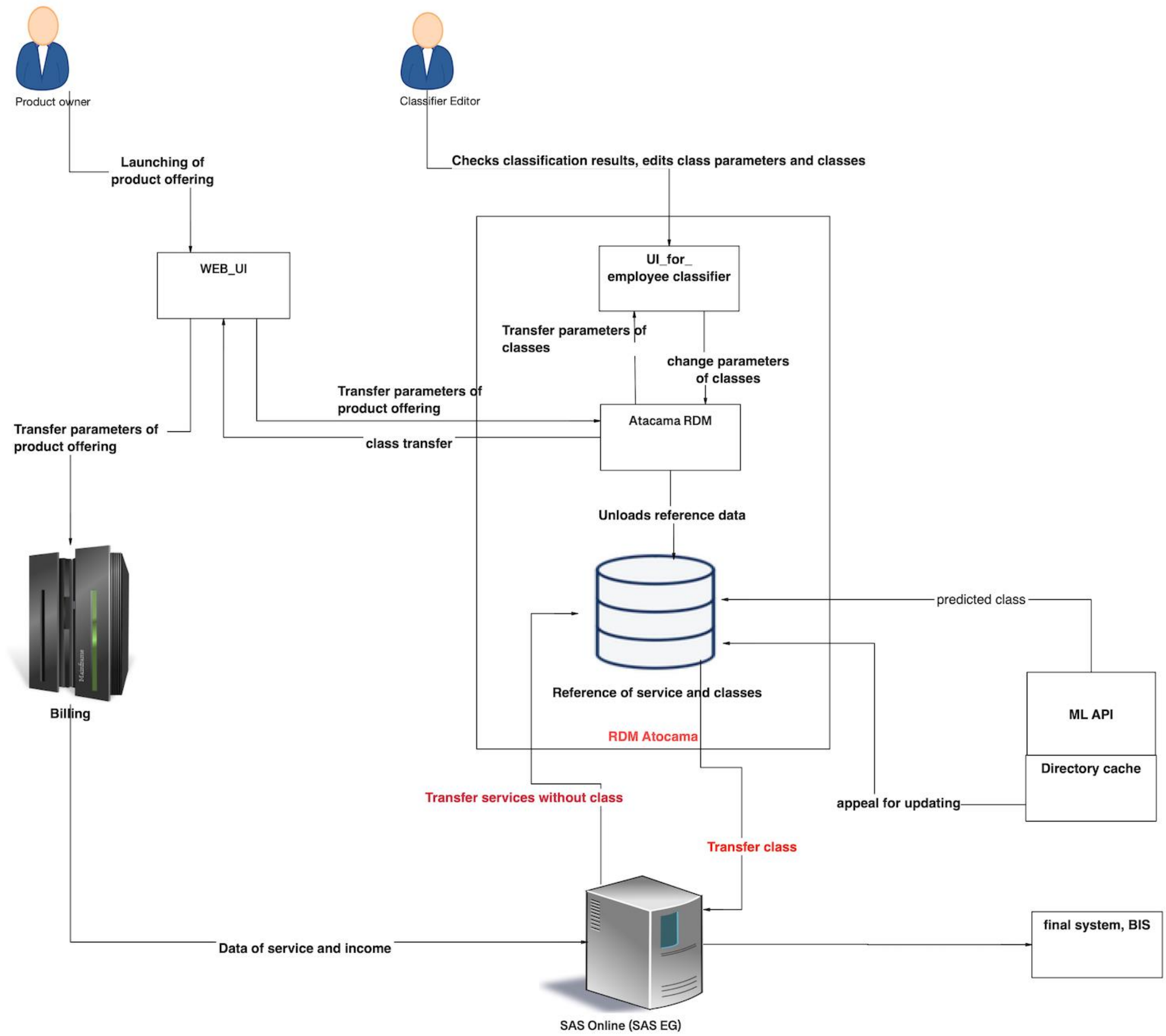

Figure 4. New architecture after loading the ML model into the company's IS loop.

In Figure 4, a new entity, the Service and Class Directory, is introduced to be developed in RDM Ataccama. An ER-model is developed (Figure 5). The ER model also specifies other tables with which integration into the Ataccama RDM Service Directory and RDM classes is required. The interface can be select REST API.

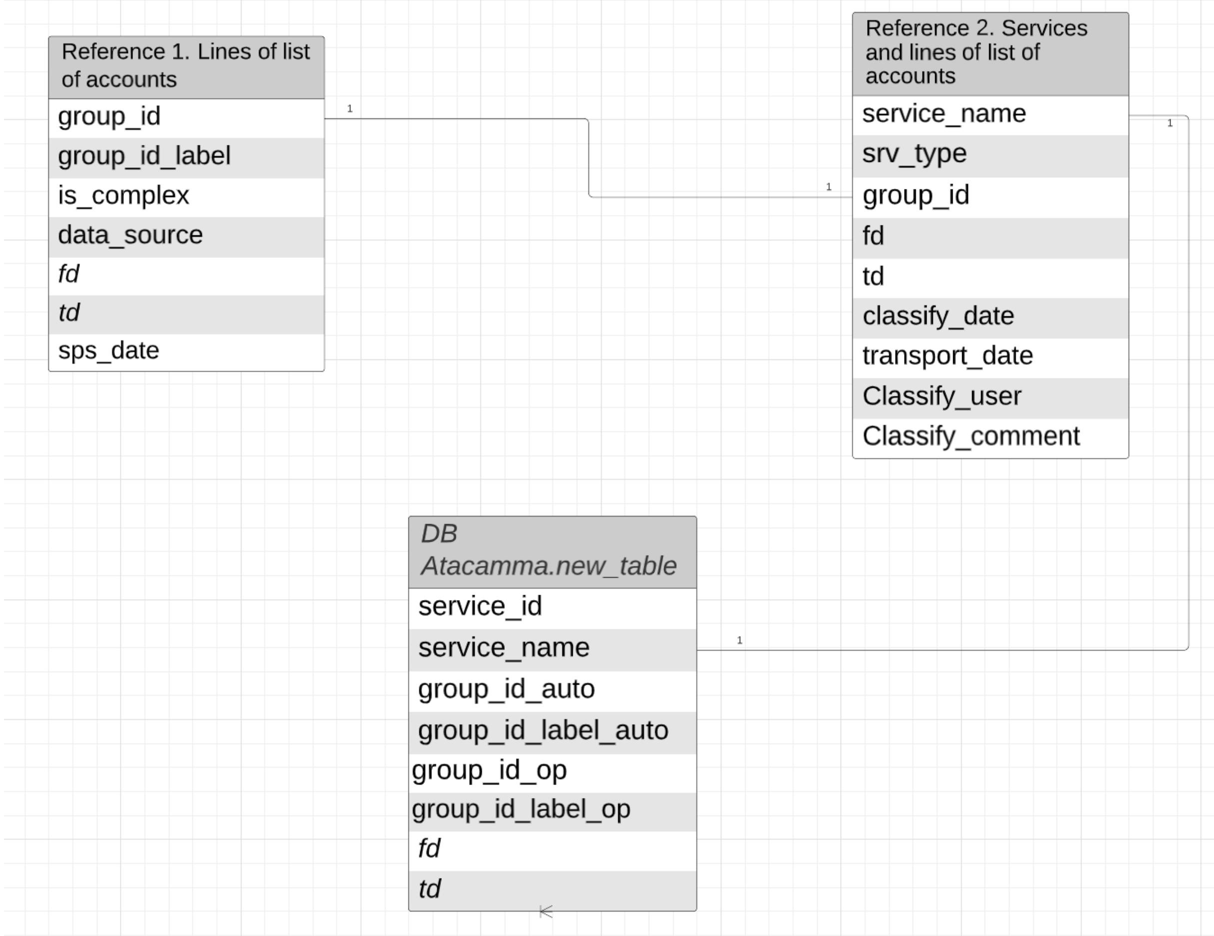

Figure 5. ER diagram of the new service and class directory and directories in account system.

Conclusion

This research considers the possibility of introducing ML algorithms and changing the existing information system to automate the business process of classifying product offers in a telecommunications company. We offer a hypothesis on developing and implementing an ML model to achieve the automation goal.

Worldwide experiences using ML models for automating business processes by telecommunications companies are analysed. Also, a review of ML algorithms used in classification problems was carried out, and algorithms were selected and employed to build the model.

Before starting modelling and evaluating the quality of the models, the current business process classification was analysed, any bottlenecks were identified, and changes were proposed. Business process models were developed in the BPMN notation of the current process (AS IS) and the process with changes (TO BE), in which changes were made, and information flows, information systems used, and users were defined. It also indicated a process that should be automated.

Next, the modelling process was carried out, and the quality of the models was evaluated according to quality metrics. The best model was selected. Model training was carried out on a prepared data set, which represents the characteristics of product offerings and their appropriate classes. The construction of a training sample can be considered as a separate task since, for its formation, it was necessary to use expert knowledge. The training used popular ML algorithms: Random Forest, Decision Tree, and XG-Boost.

After selecting the best model, a new architectural scheme was developed, considering the new elements added to the company's information system. The new elements of the information system are GUI, the development of a directory of classes, and services in RDM Ataccama. Determinations were made for the integration points of the account system directory, reporting support system, RDM Ataccama, and the new UI for the product specialist with RDM Ataccama. To make changes to the company's information system, a draft specification was developed, which includes:

- Description of functional requirements.
- Requirements for development objects.
- ER-diagram of DBMS tables (database management system) and a directory of classes and services in RDM Ataccama.
- Requirements for developing a GUI for a new interface for a product manager and creating new workspaces in the existing RDM Ataccama GUI.
- Description of user scenarios when interacting with the GUI.

In conclusion, the study solved all the tasks before it, achieving its goal and proving that the ML algorithm allows the automation of classifying services tasks. Also, the study proposes terms of reference for introducing the ML algorithm into the contour of the existing information system.

References:

1. AI Trends 2021. With machine learning, more business processes will be automated. [Accessed 21 August 2023]. https://www.aitrends.com/machine-learning/with-machine-learning-more-business-processes-will-be-automated.

2. Breiman L. 2001. Random forests. Mach Learn 45: 5–32. [Accessed 15 August 2023]. https://doi.org/10.1023/A:1010933404324.

3. Jepsen C. 2021 July 7. The truth about machine learning in enterprise software. CV Jepsen Blog. [Accessed 15 August 2023]. https://cvjepsen.blog/2021/07/07/the-truth-about-machine-learning-in-enterprise-software/.

4. Kridel, T. 2020. Telekom Malaysia leverages ML to slash contact centre queue times. [Accessed 5 August 2023]. https://www.tmforum.org/wp-content/uploads/2021/02/TM-Forum-Excellence-Awards-Case-Study-AID-Telekom-Malaysia.pdf

5. McElligott, T. (2020) Joint innovation drives China's ‘big three' towards autonomous networking. [Accessed 2021 August 1]. https://inform.tmforum.org/research-and-analysis/case-studies/joint-innovation-drives-chinas-big-three-toward-autonomous-networking.

6. Swalin A. 2018 March 13. CatBoost vs. Light GBM vs. XGBoost. Medium. [Accessed 10 July 2023]. https://towardsdatascience.com/catboost-vs-light-gbm-vs-xgboost-5f93620723db.

7. Tombes, J. 2020. Turk Telekom deploys AI and ML in network operations. [Accessed 2021 August 1]. https://www.tmforum.org/wp-content/uploads/2021/02/TM-Forum-Excellence-Awards-Case-Study-AID_Turk-Telekom.pdf

8. GSMA 2019. AI in Network Use Cases in China. [Accessed 2021 August 5]. https://www.gsma.com/greater-china/wp-content/uploads/2019/10/AI-in-Network-Use-Cases-in-China.pdf