NOVA IMS Information Management School

MDSAA

Master Degree Program in

Data Science and Advanced Analytics

Business Cases with Data Science

Case 4: Millenium – Business Process Conclusion Prediction

Carolina, Caldeira, number: 20230440

Goncalo, Caldeirinha, number: 20230469

Madalena, Figueirinhas, number: 20230436

Martyna, Kmiecik, number: 20230452

Alina, Metzger, number: 20230998

Group C

NOVA Information Management School
Instituto Superior de Estatística e Gestão de Informação

Universidade Nova de Lisboa

May, 2024

INDEX

List of Figures

List of Tables

1. EXECUTIVE SUMMARY

This report presents the development and application of predictive models for classifying task outcomes at Millennium bcp, a leading financial group in Portugal. The initiative, a collaboration between Millennium bcp and Nova IMS, leverages advanced data science techniques to enhance operational efficiency, customer satisfaction, automation accuracy, and reduced response times through accurate prediction of task execution outcomes.

Employing the CRISP-DM methodology, the project involved meticulous phases of data understanding, preparation, and modeling. The data encompassed comprehensive task execution records, user information, specific request data, and rejections, forming a robust basis for predictive analysis. Advanced machine learning techniques – including Logistic Regression, Decision Tree, Random Forest, HistGradientBoosting, and K-Nearest Neighbors (KNN) – were applied to develop models capable of accurately classifying the outcomes of customer requests.

The project's success was assessed through several metrics, primarily the F1 score, which balances precision and recall providing a comprehensive measure of model performance. These models have enabled Millennium bcp to improve task execution accuracy, reduce response times, and enhance customer service quality by predicting whether requests will be successfully executed or face rejections.

The results demonstrated varying levels of classification accuracy across different task complexities, with specific models excelling in particular contexts. For example, HistGradientBoosting emerged as the most effective model for the majority of task prefixes, demonstrating its versatility and robustness in handling different levels of data complexity. Random Forest also showed a strong performance, particularly in more complex scenarios.

To enhance the strategic impact of our predictive models, we recommend regular updates and recalibration of the models to maintain their accuracy as new data becomes available and as market conditions evolve. Furthermore, implementing a comprehensive change management plan, including employee training and support, will facilitate the adoption of these models and maximize their effectiveness.

Additionally, we suggest developing specific process optimization strategies based on the unique patterns and forecast reliability of each task type. This tailored approach will not only improve prediction accuracy but also boost overall business performance by optimizing resource allocation and enhancing customer satisfaction.

By continuing to integrate advanced analytics into operational strategies, Millennium bcp can sustain its competitive advantage in the dynamic financial services market. The successful implementation of these predictive models is expected to result in superior customer service and increased operational efficiency, solidifying Millennium bcp's market position.

2. BUSINESS NEEDS AND REQUIRED OUTCOME

2.1. BACKGROUND

Millennium bcp is a leading financial group in Portugal, known for its comprehensive range of banking services tailored for both individual and corporate clients. It operates in a highly competitive market, the financial services industry, alongside major players such as Santander Totta and Caixa Geral de Depósitos. The bank currently faces challenges with unstructured information spread across various applications, leading to inefficiencies and potential data security risks. This fragmentation results in delayed response times, inconsistent service quality, and difficulties in maintaining compliance with data protection regulations.

2.2. BUSINESS OBJECTIVES

Millennium bcp aims to efficiently manage and automate its business processes to handle high volumes of customer requests more effectively. In order to achieve this, the bank needs to enhance Business Process Management (BPM) capabilities, by integrating advanced robotic process automation (RPA) and decision engines. The project's purpose is to leverage data and predictive machine learning models to provide accurate predictions regarding the outcome of customer requests, determining whether they will be executed without rejections or face rejections. As a result, Millennium bcp aims to improve task execution by correctly identifying customer needs, route tasks to the appropriate resources, and ensure high levels of data protection and compliance.

2.3. BUSINESS SUCCESS CRITERIA

To ensure the effectiveness of the project, it is crucial to establish clear and measurable success criteria. These will serve as benchmarks for evaluating performance and guiding future business decisions. The business success criteria include:

- **Reduction in Response Times**: By streamlining processes and reducing manual intervention, the bank aims to significantly decrease the average response time for handling customer requests.
- **Operational Efficiency**: Achieving greater operational efficiency through the optimal allocation of resources, reduced processing times, and improved transparency and visibility of requests.
- **Improvement in Customer Satisfaction and Loyalty**: Providing faster and more reliable services is expected to enhance customer satisfaction. Higher satisfaction should translate into higher customer loyalty, evidenced by repeat business and positive feedback.
- **Increase in Automation Accuracy**: Enhancing the accuracy of automated processes to achieve a high percentage of correctly executed tasks without human intervention, thereby reducing errors and improving efficiency.

2.4. SITUATION ASSESSMENT

Regarding resources, a comprehensive dataset was provided, comprising data related to task execution, user information, specific request data, and rejections. A multidisciplinary team consisting of domain experts in the Business Process Management Competence Center and Automation and IT professionals from Millennium bcp, along with five data scientists from Nova IMS, collaborate closely on the project. The data is managed using Jupyter Notebook, a web-based interactive computing platform, capable of handling complex data mining tasks, along with data visualization tools.

Potential risks include technical integration issues, employee resistance to adopting new technologies, regulatory changes, and data breaches. To mitigate these risks, contingency plans include maintaining a dedicated technical support team, developing a comprehensive change management plan with training and support, ensuring project flexibility to accommodate regulatory changes, and implementing regular security audits to ensure compliance with data protection regulations.

As for terminology, it is essential that stakeholders understand the key project terms, such as evaluation metrics ‘F1 score’, ‘accuracy’, ‘precision’ and ‘recall’, as well as advanced models like ‘Random Forest’, ‘Logistic Regression’, ‘KNN’, ‘Decision Tree’ and ‘HistGradientBoost’. These concepts will be explained in the respective chapters.

While the project entails initial investment in resources and technology, its successful implementation is expected to result in superior customer service, operational efficiency and increased data protection and compliance. These improvements will also solidify Millennium bcp market position and enable it to maintain a competitive edge in the financial services industry.

2.5. Determine Data Mining goals

The technical objective is to utilize data and predictive machine learning models to accurately forecast the outcomes of customer requests, identifying whether they will be successfully executed or face rejections. This involves three main phases: data understanding and preparation; feature selection and implementation of the predictive models; results evaluation through several metrics and their interpretation. The performance of the models will be accessed through the metrics: ‘Accuracy’, ‘Precision’, ‘Recall’ and ‘F1 Score’, being ‘F1 Score’, the one determining which model is the best for each dataset.

3. METHODOLOGY

3.1. Data Understanding

In today's rapidly evolving business environment, understanding the detailed execution processes and outcomes is crucial for making informed decisions. This section delves into Millenium’s datasets concerning task executions, user information, specific requests, and rejections, which together form a comprehensive foundation for predictive analysis and process optimization.

User information dataset offers comprehensive profiles of individuals, including their roles, years of experience, and departmental associations. Such data is instrumental in deciphering workforce dynamics and identifying prospects for targeted training and development. These datasets include specifics about users, such as their IDs, gender, role, birth year, and more.

Analysis reveals that a majority of users cluster around a few key role IDs, with role 21 standing out for its prevalence. This implies that a subset of roles exerts considerable influence within the organization's hierarchy. Figure 1 shows that over the past four years (2020-2024), there has been a noticeable increase in the addition of new members to various organizational departments, which could indicate recent changes or growth in the organization's structure.

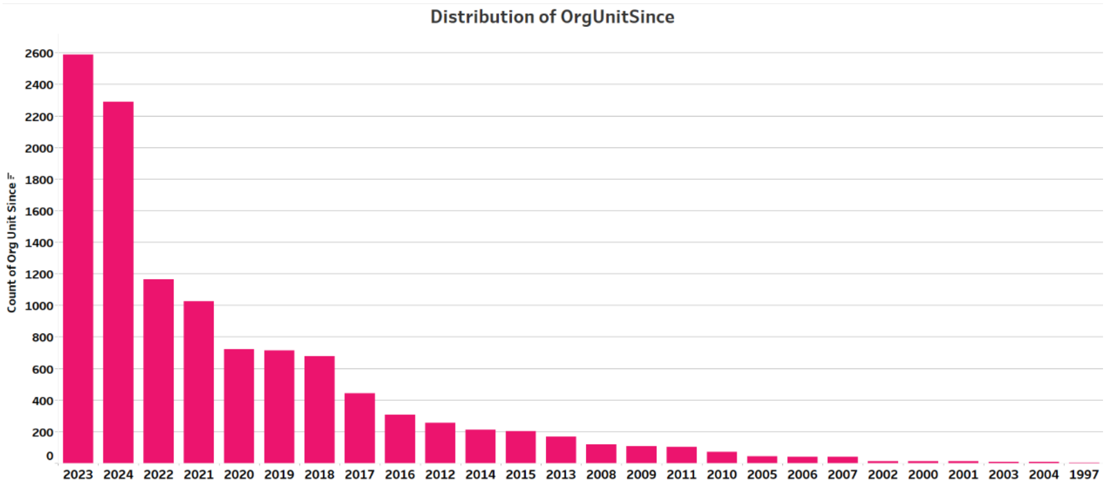

Figure 1: Distribution of the Years when Employment with Organizational Unit Started

Regarding gender balance, the user base is predominantly male, with a notable percentage of users holding managerial positions. Additionally, only 36% of users come from outsourcing, indicating a preference for utilizing internal talent.

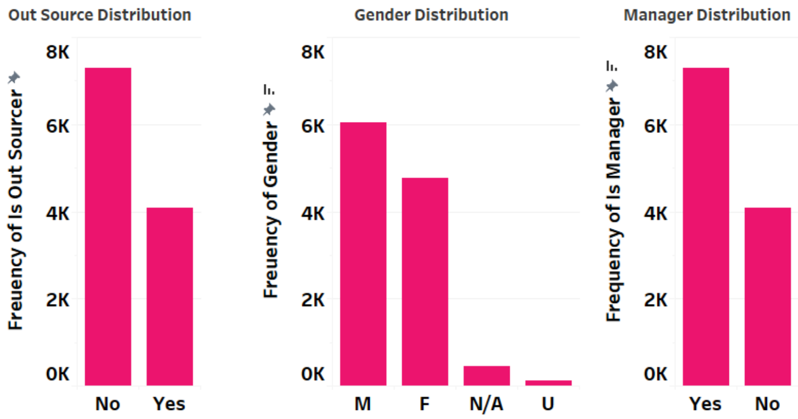

Figure 2: Distribution of Outsourcers, Gender and Managers

The task execution dataset contains detailed information on each task performed, including the task ID, request identifier, task initiation and completion dates, and the department overseeing the task. This data is crucial for evaluating operational efficiency. An examination of task execution durations can uncover patterns in workflow efficiency, highlighting areas where enhancements can be implemented. Tasks that consistently exceed expected durations may signal underlying process inefficiencies or resource constraints.

Further analysis of task execution data uncovers additional insights. The prevalence of activity IDs suggests that certain tasks are more commonly encountered or critical within the organization. Activity ID 102 emerges as the most frequent, followed by IDs 104, 100, 103, 107, 101, 105, 108, and 106, hinting at their importance to operations. The categorization of task types reveals that execution tasks predominate, significantly outpacing initial requests, final tasks, and responses to rejections. This underlines that the core activities center around task execution rather than initiating requests or finalizing responses.

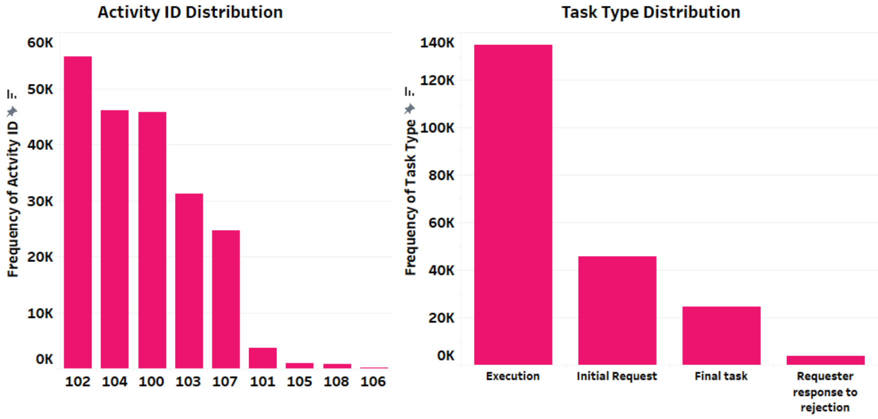

Figure 3: Distribution of Activity ID and Task Type

In examining the challenges presented by task execution data, the concept of outliers becomes particularly significant. Outliers are data points that deviate markedly from other observations. For instance, in the data, most requests that conclude with the application action ID 8888 finish with activity ID 104. However, instances where activity IDs 107 and 102 occur with the same application action ID are rare, suggesting they are outliers. This anomaly might be indicative of different processes being followed in the past, or it could signify exceptionally special requests that do not follow the usual patterns. Identifying such outliers is crucial as they can reveal underlying issues or changes in the process that need further investigation.

The **specific requests** dataset logs detailed information on the types of requests made, while the **rejections** dataset holds information on tasks or requests that were declined or sent back for various reasons.

It is important to note that all data is masked to ensure privacy and prevent data leakage. Comprehending the vital need to safeguard confidential information and anonymizing all data helps reduce risks connected to data breaches and maintain the trust entrusted by users and stakeholders. This strategy permits the examination and improvement of processes without jeopardizing personal privacy or the overall integrity of Millenium and its clientele.

3.2. DATA PREPARATION

Initial Preprocessing

After data understanding, it is crucial to tackle the inconsistencies and extract the important information for the modeling of the problem. As the data has a complex structure, the strategy implemented is constructed of carefully considered techniques to maximize the results' quality and interpretability.

The first steps included checking for duplicated data and the data types which were all appropriately dealt with. The data comprised 4 different tables: Task Execution Data, User Information, Specific Request Data, and Rejections. To facilitate the preprocessing stage, the first three datasets were merged. The fourth one, however included duplicated information and no contextualization attached to it. After the exploration was carefully conducted and no direct relation of that information to the other data was found, the Rejections dataset was excluded. This decision was also made to maintain clear, interpretable results and minimize incorrect data processing.

Following this step, new variables related to the duration of the tasks and requests, as well as the day and month of each event were added extracting important information from date/time-related columns.

Some inconsistencies regarding changes in the process were found, namely related to BPM Application Action codes, as there were two on the earlier records that did not appear on the most recent ones. Additionally, there were two dates where around 10,000 requests were closed administratively, which could indicate that the process was different before those dates. However, not all active requests were created before those two dates were closed and some continued the process afterwards. Since there was no clear date as to when the process changed and in the following steps a significant amount of data was deleted, including outliers, it was decided to not delete data based on a temporal threshold.

Following this step, a column indicating the outcome of the process was added. The possible outcomes are as stated in the process diagram (Figure 4): 'Request Finished', 'Request Canceled', 'Others' (Closed Administratively, Requester Rejects, Accounting Impact), and 'Closed Administratively'. These outcomes are associated with specific BPM Application Action codes as shown in the diagram, as well as specific Activity IDs. However, some exceptions were found, where a request was in a certain activity and the outcome was unexpected. Since the Activity IDs would be used for prediction, these exceptions were removed, to bring light to the clear connection between the outcomes and the activities. For example, for a request to have the outcome 'Request Canceled', it would need to go through Activity 101. In case there was a request that went through Activity 106 and had the outcome 'Request Canceled', this was removed, as it represents an outlier and could confuse the relationships the model will identify between variables.[1]

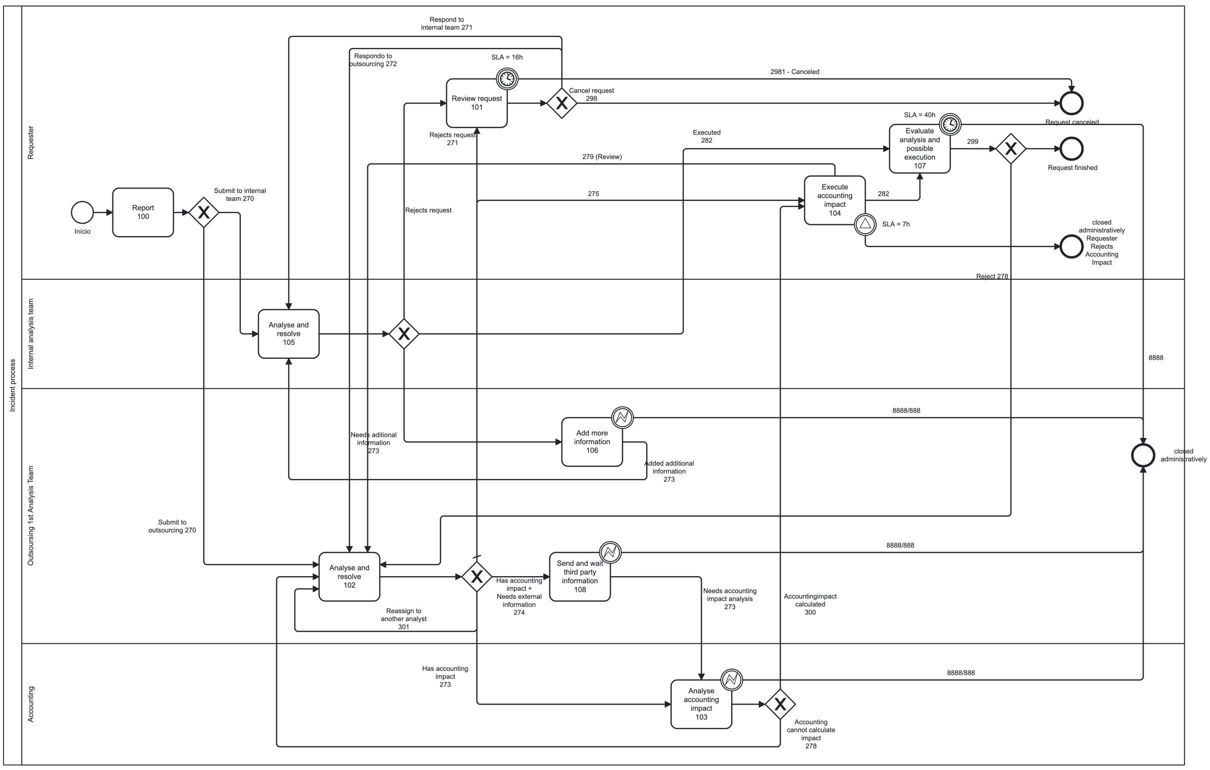

Figure 4: Process Diagram

[1] This example is merely illustrative and does not represent an actual request.

Prefix Tracing and Other Transformations

As the focus of this proposal is Predictive Process Monitoring, and the objective is to predict the outcome of a request at certain activities (100, 102, 105), the data had to be transformed.

This transformation organized the data in a way that each row corresponded to a request and all the sequential information of a request was contained within that row. This way, all the changes of information throughout a request are stored and will be used as input to obtain the predicted outcome of that row.

Following this change it was important to identify the 'Prefixes' (illustrated in Figure 5). This is a technical concept that will be mentioned in this report. Since each row in the data was transformed to be a sequence of different events (the request goes through different activities, each of these moments can be considered an event) and the prediction might be done at any point (because the request can be at activities 100, 102, 105 at any point in the process), it is important to create models for each of these events. Looking at a practical example, if a request went through the following activities - 100, 102, 101, 102, 108, outcome – the model should be able to predict the outcome on the first, second, and fourth moment. For this reason, nine datasets were created one for each Prefix from the first to the ninth, making sure that all past information is stored. The following schema demonstrated this transformation. This is what allowed the data to be used for predicting the outcome.

Task	Request Identifier	Activity	Task Executor	Other features
1	1	100	1046	...
2	1	102	306	...
3	2	100	4867	....
4	2	105	1345	

Request Identifier	Activity 1	Task Executor 1	(Other features 1)	Activity 2	Task Executor 2	(Other features 2)
1	100	1046	...	102	306	...
2	100	4857	...	105	1345	...

Prefix 1

Request Identifier	Activity 1	Task Executor 1	(Other features 1)
1	100	1046	...
2	100	4857	...

Prefix 2

Request Identifier	Activity 1	Task Executor 1	(Other features 1)	Activity 2	Task Executor 2	(Other features 2)
1	100	1046	...	102	306	...
2	100	4857	...	105	1345	...

Figure 5: Prefix Trace Representation

Prefix 1, has information for the first activity of all requests, Prefix 2 has all the all the information for the first activity of all requests and information on the second activity of all requests, and so on, until the 9th Prefix has all the information of past activities and a selection for the 9th activity. It was decided that the prediction was made at the moment the request entered the last activity, so no future information was included. This restriction reflects how in real life, when predicting the outcome at a certain point, there would be no information after that moment, which means there cannot be future information after the last occurrence of the important actitvites.

Although some requests have more than 9 activities, these were disregarded because they were a very small percentage of data.

Finally, to finish this transformation, all rows of data in all Prefixes where the last activity was not 100, 102, and 105 were deleted, to ensure that the models predict the outcome when the request is on exactly these three activities.

Following this transformation, the columns that were no longer necessary were dropped (since new ones with the important information were already created prior).

Finally, the data was split into two data sets: training and validation, one to train the model and the other to validate results and evaluate its performance.

The following steps were taken to prepare the data to be used for the specific models (which will be described in the following chapter 3.3). At this stage, missing values were filled for both sets, based on the information from the training (this is very important, to make sure that when the models do their prediction, it is done on new, unseen data). Lastly, since the data is a mixture of categorical and numerical, the numerical data was scaled, and the categorical data was encoded. Scaling numerical is a method to adjust the range and center of the data, ensuring consistency across the features. Additionally, it improves model performance.[2] Encoding categorical features changes the information into numbers, as it is necessary for some models that can only process numerical formats.[3]

Selection of Important Features

The last step before training the models is to select the important features for each of the Prefixes. As there were many variables for each of the Prefixes, the first selection method was based on the variance of the variable. All variables with variance lower than a specified threshold were not considered because that indicates that the values of that feature do not differ significantly.

Following that, two other techniques were implemented: *Recursive Feature Elimination* and *LassoCV* with cross-validation. *Recursive Feature Elimination*[4], builds a model with all features, sequentially removing the least important each repetition, until the defined number of features is reached. *LassoCV*[5] is indicated for when there is a considerable number of features but only a few are relevant for the prediction.

The following features were identified as the most relevant for prediction: Field 203, Field 1604, Field 3491, Task Executer preceding the final activity, Arrival Day of the final activity, Role of the task executer and BPM Application Action code.

For more information on data preprocessing, please refer to the source code.

3.3. Modeling

Four potential outcomes – 'Request Finished', 'Request Canceled', 'Closed Administratively', and 'Others' – were examined. The category 'Others' accounts for the largest portion at 40.56%. Following this, 'Closed Administratively' outcomes, representing tasks closed without completion, constitute

[2] Cf. Baruah, I. D. (2023).
[3] Cf. Analytics Vidhya (2022).
[4] Cf. Brownlee, J. (2020).
[5] Cf. Agrawal, S. (2023).

34.14% of the total. The ‘Request Finished’ category, indicative of successfully completed tasks, makes up 20.75%. Lastly, the ‘Request Canceled’ outcome, representing tasks that were not successfully completed and subsequently canceled, comprises 4.55% of the total outcomes.

To enhance analysis consistency, ‘Closed Administratively’ and ‘Others’ were managed identically as canceled outcomes. This strategy fosters uniformity and simplifies the classification process, thereby improving the predictability and interpretability of predictive models.

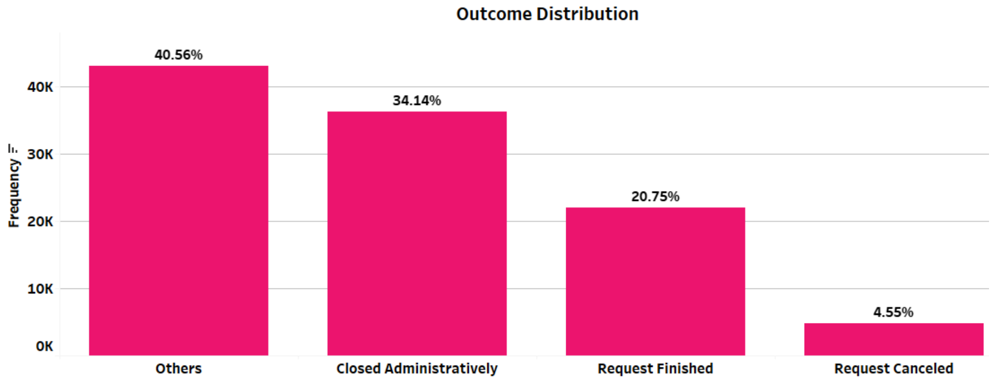

Figure 6: Distribution of Outcomes

Furthermore, to analyze the outcomes of task executions, a variety of machine learning models were employed, and their efficacy was assessed through both training and validation phases. A broad spectrum of models was chosen, each recognized for its distinctive capabilities in managing different data patterns and relational structures:

- **Logistic Regression**: This method is easy to understand and use, providing estimates on the likelihood of different outcomes.
- **Decision Tree**: This model splits data into branches to help make decisions, making it simple to visualize and understand the decision-making process.
- **Random Forest**: This approach combines many decision trees to improve accuracy and reliability in predictions.
- **HistGradientBoosting**: This method builds a series of decision trees one by one, learning from previous mistakes to capture complex patterns effectively.
- **K-Nearest Neighbors (KNN)**: This technique categorizes data points based on the most common outcome among the nearest data points, making it easy to group similar instances together.

To maximize the models' performance, an advanced method, called *GridSearch*, was implemented to determine the optimal parameters for each model. This methodical strategy explored various parameter combinations to ascertain the most beneficial setup for each model.

3.4. Evaluation

When evaluating the performance of different models for predicting process outcomes based on different task complexities (denoted by Prefixes), selecting the model that balances accuracy with the ability to generalise on unseen data is crucial. The evaluation focuses primarily on the F1 score, however different evaluation metrics were also taken into account:

- **Accuracy** measures the percentage of all predictions that are correct.

- **Precision** indicates how many of the items the model labeled as a particular category are actually in that category.
- **Recall** reflects how many of the actual items in a category the model correctly identifies.
- **F1 Score** combines recall and precision offering a harmonious measure of model performance, particularly valuable when it is equally important to avoid incorrect predictions and to capture as many relevant cases as possible.

The detailed effectiveness of the five different models used for every request complexity is presented in the table below:

Dataset	Metric	Logistic Regression	Decision Tree	Random Forest	HistGradientBoost	KNN
Prefix 1	**F1 Score**	0.4714	**0.5011**	0.4999	0.5010	0.4592
	Accuracy	0.5065	0.5220	0.5210	0.5217	0.4820
	Precision	0.5486	0.5716	0.5693	0.5715	0.4687
	Recall	0.5065	0.5220	0.5210	0.5217	0.4820
Prefix 2	**F1 Score**	0.4668	0.4956	0.5023	**0.5075**	0.4714
	Accuracy	0.5037	0.5214	0.5319	0.5359	0.4946
	Precision	0.5462	0.5635	0.5881	0.5800	0.483
	Recall	0.5037	0.5214	0.5319	0.5359	0.4946
Prefix 3	**F1 Score**	0.3693	0.4604	0.4673	**0.4731**	0.4497
	Accuracy	0.4698	0.5000	0.5155	0.5009	0.4672
	Precision	0.3091	0.4580	0.4752	0.4608	0.4366
	Recall	0.4698	0.5000	0.5155	0.5009	0.4672
Prefix 4	**F1 Score**	0.2791	0.4405	0.4166	0.4286	**0.4413**
	Accuracy	0.4062	0.4732	0.4643	0.4721	0.4732
	Precision	0.2400	0.4477	0.4364	0.4598	0.4477
	Recall	0.4062	0.4732	0.4643	0.4721	0.4732
Prefix 5	**F1 Score**	0.3647	0.4651	0.4537	**0.5357**	0.517
	Accuracy	0.4767	0.4990	0.4949	0.5578	0.5416
	Precision	0.3192	0.4640	0.4622	0.5577	0.519
	Recall	0.4767	0.4990	0.4949	0.5578	0.5416
Prefix 6	**F1 Score**	0.3877	0.3900	0.4250	**0.4515**	0.4415
	Accuracy	0.4452	0.4110	0.4692	0.4623	0.4555
	Precision	0.4064	0.3839	0.4308	0.4464	0.4397
	Recall	0.4452	0.4110	0.4692	0.4623	0.4555
Prefix 7	**F1 Score**	0.3948	0.4349	**0.4834**	0.4511	0.4042
	Accuracy	0.4464	0.4673	0.5446	0.4732	0.4286
	Precision	0.3953	0.4466	0.5387	0.4363	0.4438
	Recall	0.4464	0.4464	0.5446	0.4732	0.4286
Prefix 8	**F1 Score**	0.3634	0.2895	0.3643	**0.4665**	0.3762
	Accuracy	0.3913	0.2899	0.4058	0.4783	0.4638
	Precision	0.4539	0.2987	0.3607	0.4864	0.3634
	Recall	0.3913	0.2899	0.4058	0.4783	0.4638
Prefix 9	**F1 Score**	0.3226	0.3105	**0.3782**	0.3345	0.3779
	Accuracy	0.3333	0.3000	0.4333	0.3333	0.4000
	Precision	0.3328	0.3247	0.3714	0.4067	0.4362
	Recall	0.3333	0.3000	0.4333	0.3333	0.4000

Table 1: Evaluation Metrics for the Models

All of the models tested were implemented with the best parameters for every case scenario based on the results of *GridSearch*, pointing to highest performing attributes. Nevertheless, the obtained evaluation metrics differ significantly between the models and datasets.

The *HistGradientBoost* model proved to be superior not only in the less complex but also in the more complex scenarios, classifying the outcomes of the cases finishing with activity 100, 102 or 105 the best for the highest number of requests, that is for the ones consisting of 2, 3, 5, 6 and 8 activities. That demonstrates model's versatility and robustness in handling various levels of data complexity and task interactions, thus making it a strong tool for predicting results in these situations.

Random Forest performed the best among the models for two Prefixes – 7 and 9. This makes it the second-best performing model overall. It excels in handling the most complex datasets, as these Prefixes involve a significant number of features selected through feature selection. That means that the Random Forest model's strength lies in its ability to manage high-dimensional data and capture intricate patterns and interactions within the data. Therefore, for datasets with many features and high complexity, Random Forest proves to be an excellent choice, providing robust and reliable performance.

Both the *Decision Tree* and the *KNN* model performed the best amongst the others for two different requests lenghts. The first one obtained higher F1 scores than other techniques for cases composed of one task, indicating its strength in handling less complex scenarios with smaller datasets. On the other hand, the KNN model excelled in cases involving four activities, showing its capability in slightly more complex but still manageable situations. Therefore, for less complex scenarios and smaller datasets, using Decision Tree and KNN models can provide better performance and more reliable predictions compared to other techniques, which is why in these particular scenarios, the choice of these algorithms would be advisable.

The evaluation of the results showed that *Logistic Regression* proved to be the least effective in every case, as they never obtained the highest F1 scores among other models tested.

Overall, the combined evaluation scores for the best performing models were as follows: accuracy and weighted recall - 52.91%, weighted precision - 55.51%, and the F1 score - 50.25%. These results, while leaving room for improvement, indicate a respectable level of performance considering the complexity of the data and the challenges involved.

In general, for processes encompassing more than 5 tasks, a general trend of performance degradation across all models can be observed. This phenomenon is likely due to the sparse and lower amount of data available for training the models on these high-complexity tasks, leading also to overfitting issues. Overfitting occurs when models are too closely tailored to the training data, limiting their ability to generalize to new, unseen situations. *HistGradientBoosting* still performs relatively better, but the drop in F1 scores indicates the need for more nuanced data collection and model training strategies for higher complexities of requests.

To address this problem, expanding the dataset to include more examples of higher complexity tasks will be critical. Additionally, exploring more sophisticated machine learning techniques that can better generalize from limited data might prove beneficial.

Moreover, none of the models, including *HistGradientBoost*, performed well in predicting the outcome of requests being cancelled, especially for the final activity 100. This is likely due to the limited number of such cases, which prevents the models from being well-trained in this specific scenario. Additionally, for activity 105, there were many predicted outcomes classified as 'Others', even though there were no actual values labeled as such for this scenario. These issues indicate areas where the model's performance could be improved, with collecting more requests with such characteristics. Requests closed administratively are predicted the most successfully, regardless of the end activity of the request. This consistent performance across different end activities highlights the models' effectiveness in recognizing patterns associated with administratively closed requests. The reliable prediction of this outcome, suggests that the features and interactions within the dataset are well-captured and leveraged by the model, ensuring accurate classification for administratively closed cases.

Overall, based on the F1 Score, which consists of two evaluation metrics precision and recall, *HistGradientBoost* model may be considered the best fit – not only for the shorter, but also for the longer requests. Nevertheless, there is a clear need to refine the approach for higher-complexity scenarios, with broadening the amount of less frequent cases to train the algorithms.

4. RESULTS EVALUATION

The models developed aim to predict the outcomes of customer requests accurately, directly aligning with Millennium business objectives. The primary goals include reducing response time, increasing operational efficiency, improving customer satisfaction, and enhancing automation accuracy.

The robust performance of the models in predicting outcomes for activities 100, 102 and 105, even with varying task complexities, shows significant promise in meeting the business objectives. It is noticeable that the predictive models have the potential to significantly improve task execution accuracy. However, the full measurement of these business success criteria can only be done after implementation in a real environment. Deploying the models within Millennium bcp's actual BPM workflows, and continuously monitoring their performance in live settings, will provide actual evidence of their effectiveness.

By predicting outcomes accurately, the models enable proactive management of customer requests, potentially leading to reduced processing times. Correctly anticipating whether a request will be successfully executed, or face rejections, allows for faster routing and handling, thereby decreasing overall response times. It also ensures that tasks are routed to the appropriate resources, minimizing delays and enhancing process transparency. Faster and more accurate processing of customers' requests is likely to enhance customer satisfaction. Reliable service and reduced waiting times contribute to a better customer experience, fostering loyalty and repeat business. Although this project's results indicate potential improvements, actual customer feedback and loyalty metrics post-implementation are necessary to provide concrete evidence of improvement in customer satisfaction and loyalty. The considerable high accuracy shown by the models demonstrates its effectiveness in automating the decision-making process, which reduces human error and ensures consistent quality in task execution.

The following bar plots compare actual and predicted outcomes for activities 100, 102 and 105.

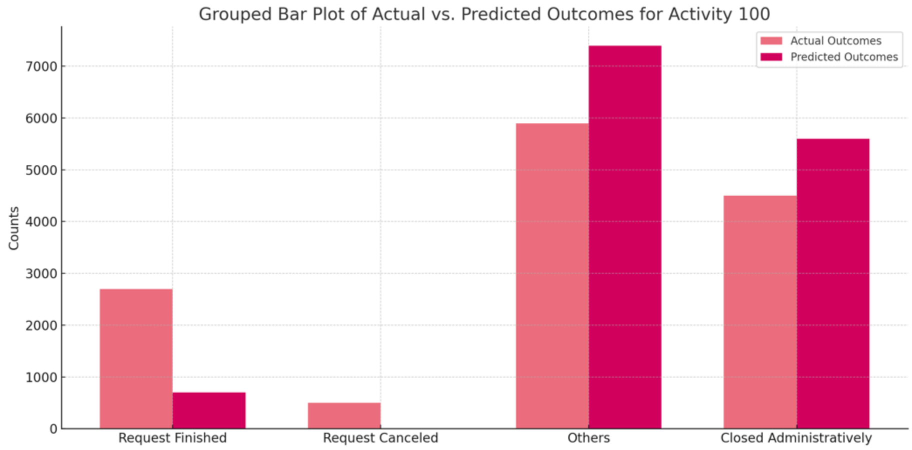

Figure 7: Outcomes of Activity 100

Regarding activity 100, the model predicts considerably well all the outcomes instead of 'Requests Canceled', that the model seems to not predict. A reason for this is the low number of requests where this happens, which makes it harder for the model to train on and then predict. According to the data, activity 100 is followed mainly by the outcome 'Others' which refers to 'Closed Administratively', highly an outcome on this context, 'Requester Rejects' and 'Accounting Impact'.

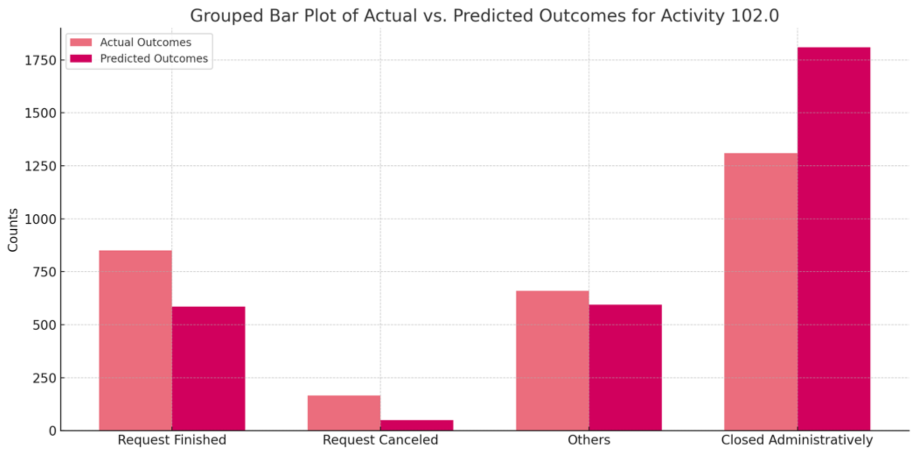

Figure 8: Outcomes of Activity 102

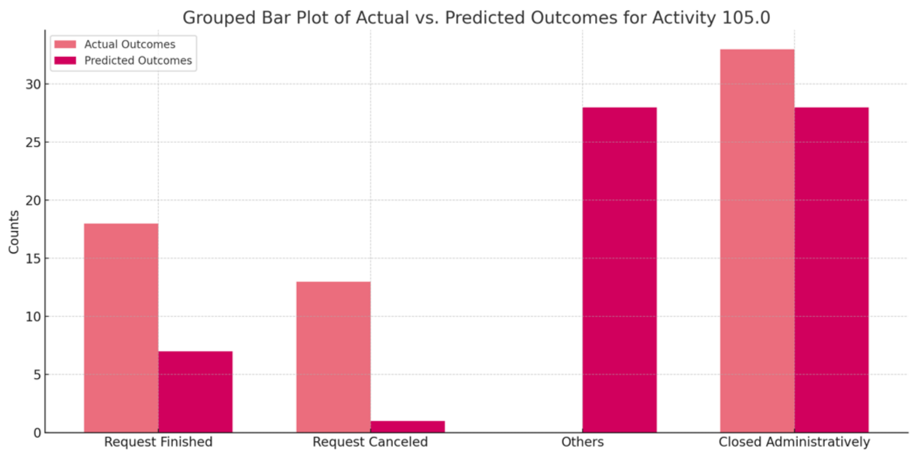

Figure 9: Outcomes of Activity 105

Regarding activities 102 and 105, the predictions are more accurate for all outcomes in activity 102 instead of 105. In outcome prediction regarding activity 105, the model predicts several 'Others' outcomes while there are no actual ones, this suggests potential areas for model improvement that need to be followed by time and more computational resources. 'Closed Administratively' is the outcome more common in both activities and the outcome better predicted in general. These two characteristics are connected, since the models predict better with more data, a crucial aspect for accurate predictions that is missing in this context, leaving room for improvement.

5. DEPLOYMENT AND MAINTENANCE PLANS

Deploying the developed model into production and maintaining its optimal performance requires a comprehensive strategy to ensure seamless integration and long-term effectiveness while maintaining high standards of data protection and compliance – a priority for Millenium. The following deployment plan outlines the necessary steps for successful implementation:

Infrastructure Setup:

- **Cloud or On-Premises Deployment**: Deciding whether the model will be deployed on a cloud-based platform or on-premises, considering factors like scalability, security, and cost.
- **Server configuration**: Configuring servers to handle the model's computational requirements, ensuring they are optimized for load balancing and fault tolerance.

System Integration:

- **Integration with Existing System**: Integrating the predictive model into Millennium bcp's existing CRM-similar system, which was preselected for deployment. This involves creating APIs (Application Programming Interfaces) that allow the CRM to communicate with the model.
- **Testing in Sandbox Environment**: Deploying the model in a sandbox environment that mirrors the production environment to conduct thorough testing. This includes testing the entire data flow and system integration to ensure all components work together as expected.

Data Pipeline:

- **Data Flow Integration**: Ensureing seamless data flow between the CRM-like system and the predictive model. This includes setting up extraction, transformation and loading processes to clean and prepare data.
- **Real-time Data Processing**: Implement mechanisms to process incoming requests in real-time, providing immediate predictions.

Security and Compliance:

- **Data Security Measures**: Implementing robust security measures to protect sensitive customer information. This includes encryption for data at rest and in transit to protect sensitive client information as well as access controls to ensure only authorized personnel can access the data and the predictive model.
- **Regulatory Compliance**: Ensuring the deployment complies with relevant data protection regulations (e.g., GDPR) and maintain detailed audit trails to track access and modifications.

Training and Support:

- **Employee Training**: Conducting training sessions for employees who will interact with the CRM-type system and use the predictive model's outputs.
- **User Documentation**: Providing detailed documentation for users to facilitate usage and troubleshooting.

Following deployment, ongoing monitoring and maintenance of the predictive model will be essential to ensure its continued effectiveness and relevance. The strategy includes, but is not limited to, the measures listed below that address various aspects of performance, security and compliance:

- **Regular Model Retraining**: Schedule periodic retraining of the model with new data to ensure it stays up-to-date and accurate.
- **Performance Monitoring**: Continuously monitor model performance metrics (accuracy, precision, recall, F1 score) to detect any degradation over time.
- **Data Quality Checks**: Implement automated checks to ensure the quality and integrity of incoming data.
- **Anomaly Detection**: Use anomaly detection techniques to identify and address any unusual patterns or errors in the data or predictions.
- **Compliance Reviews**: Regularly review compliance with regulations and standards to ensure ongoing adherence.
- **Security Audits**: Conduct regular security audits to ensure compliance with data protection regulations and to safeguard sensitive customer information.
- **Feedback Loop**: Incorporate user feedback to refine and improve the model's predictions.
- **System Health Monitoring**: Monitor the health of the infrastructure, including server performance, uptime, and resource utilization.
- **Error Logging and Alerts**: Implement comprehensive error logging and set up alerts for any system failures or issues that require immediate attention.
- **Model Versioning**: Maintain version control for the model to track changes and ensure rollback capabilities if needed.
- **Documentation Updates**: Keep documentation up-to-date with any changes to the model, data pipelines, or system integrations.

Regarding the personnel involved in the deployment and maintenance process of Millenium's model predicting business process outcomes, the following table outlines the key stakeholders along with their respective roles and responsibilities.

Personnel	Role	Responsibilities
BPM Competence Center and Automation Experts	Subject Matter Experts	Provide domain expertise and ensure the model aligns with business process management and automation objectives
Data Scientists	Model Development and Tuning	Develop, fine-tune, and monitor the predictive model
IT Infrasturcture Team	Infrastructure Management	Set up and maintain servers and cloud resources, ensure system scalability and availability
Integration Specialists	System Integration	Develop and manage APIs for data exchange, ensure seamless integration between CRM-similar system and predictive model
Security Analysts	Data Protection and Compliance	Conduct security audits and ensure compliance with data protection regulations
Support Team	Technical Support	Provide technical support during and after deployment, handle user issues and queries

Table 2: Roles and Responsibilities

To ensure a systematic deployment and maintenance of predictive model, a well- defined timeline is essential. Below is a brief overview of the proposed schedule.

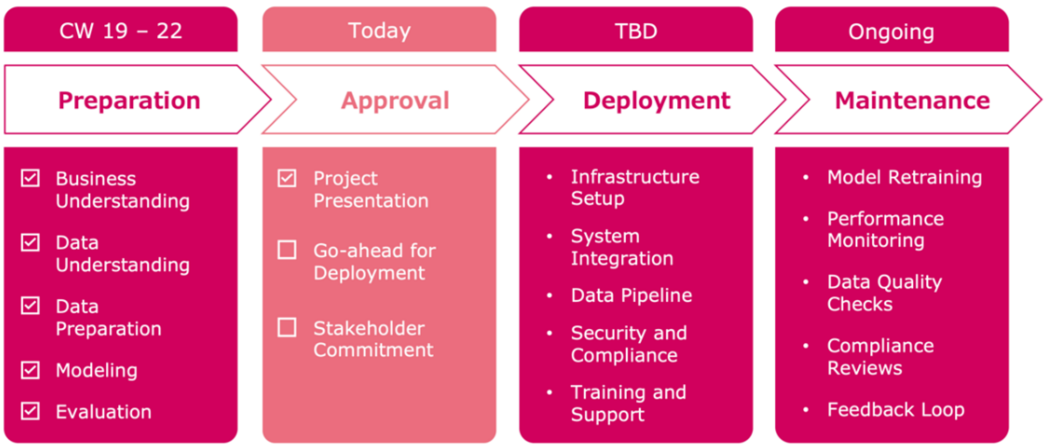

Figure 10: Project Plan

6. CONCLUSIONS

The development of the presented predictive machine learning model to forecast the outcomes of customer requests at Millennium bcp marks a significant step towards refining operational efficiencies, reinforcing Millennium's position as a competitive player in the financial services market. This model not only demonstrates promising results for earlier stages of the requests but also reflects the company's commitment to the use of data science for strategic advantage. A successful implementation of this proposal could lead to quicker response times, higher costumer satisfaction, as well as more refined internal process operations.

The focus on continuous improvements and stakeholders commitment is crucial to garantee the success of this project. Continuous model training and a focus on enhanced model generalization, while reducing overfitting should be prioritized in order to improve results. Aditionally, incorporating more sophisticated machine learning techniques and process mining tools, leveraging the companies software and technological capabilities, could offer deeper insights into the current state of the process, facilitating more informed decision-making and optimization.

6.1. Considerations for Model Improvement

To further enhance the model predicting business process outcomes, the following recommendations are proposed:

- **Enhancing Model Generalization**: Addressing the observed overfitting in processes with more than five tasks is crucial. This highlights the challenge of limited training data for complex scenarios. To counter this, it is critical to expand the dataset including more examples of higher complexity. This broader dataset will help the model learn more robust patterns, improving its performance in real-world scenarios without relying too heavily on specific training instances.
- **Leveraging Process Mining Tools**: Implementing process mining solutions (e.g., Celonis) can provide valuable insights into existing workflows and process executions. These tools analyze operational data to visualize process flows, identify bottlenecks, and uncover inefficiencies. By integrating process mining with predictive modeling, Millennium bcp can gain a comprehensive understanding of its business processes, enabling data-driven decision-making and targeted optimizations. This holistic approach ensures that predictive models are built on a solid foundation of process understanding, enhancing their accuracy and relevance to real-world operations.
- **Collaborating with Domain Experts**: Engaging more closely with experts from Millennium bcp's BPM Competence Center would be helpful. Their insights into domain-specific factors influencing business process outcomes can refine model assumptions. They can, for example, provide valuable context regarding process evolution, such as when processes changed or new elements were introduced. This contextual information enriches the model training process, ensuring that predictive models adapt to dynamic business environments more effectively.
- **Interpretability and Explainability**: Improving model interpretability and explainability is key to helping stakeholders understand model predictions and trust the outputs. This ensures transparency in decision-making processes and fosters confidence in the predictive model's reliability. Achieving this goal requires employing transparent modeling techniques, such as decision trees or linear models. However, it is essential to train these models effectively to ensure they provide accurate results that can be reliably interpreted, making them preferable choices for enhancing model transparency and stakeholder understanding.

Incorporating these consideration will not only enhance the model's predictive capabilities but also empower Millennium bcp to make more informed decisions, ultimately driving greater efficiency and success in business process management.

7. REFERENCES

Agrawal, S. (2023). Feature Selection Using Lasso Regression. Retrieved from *Medium*: https://medium.com/@agrawalsam1997/feature-selection-using-lasso-regression-10f49c973f08.

Baruah, I. D. (2023). All you need to know about encoding techniques! Retrieved from Medium: https://medium.com/anolytics/all-you-need-to-know-about-encoding-techniques-b3a0af68338b.

Brownlee, J. (2020). Data Preparation.

Feature Scaling Techniques in Python – A Complete Guide. (n.d.). Retrieved from Analytics Vidhya: https://www.analyticsvidhya.com/blog/2021/05/feature-scaling-techniques-in-python-a-complete-guide/.