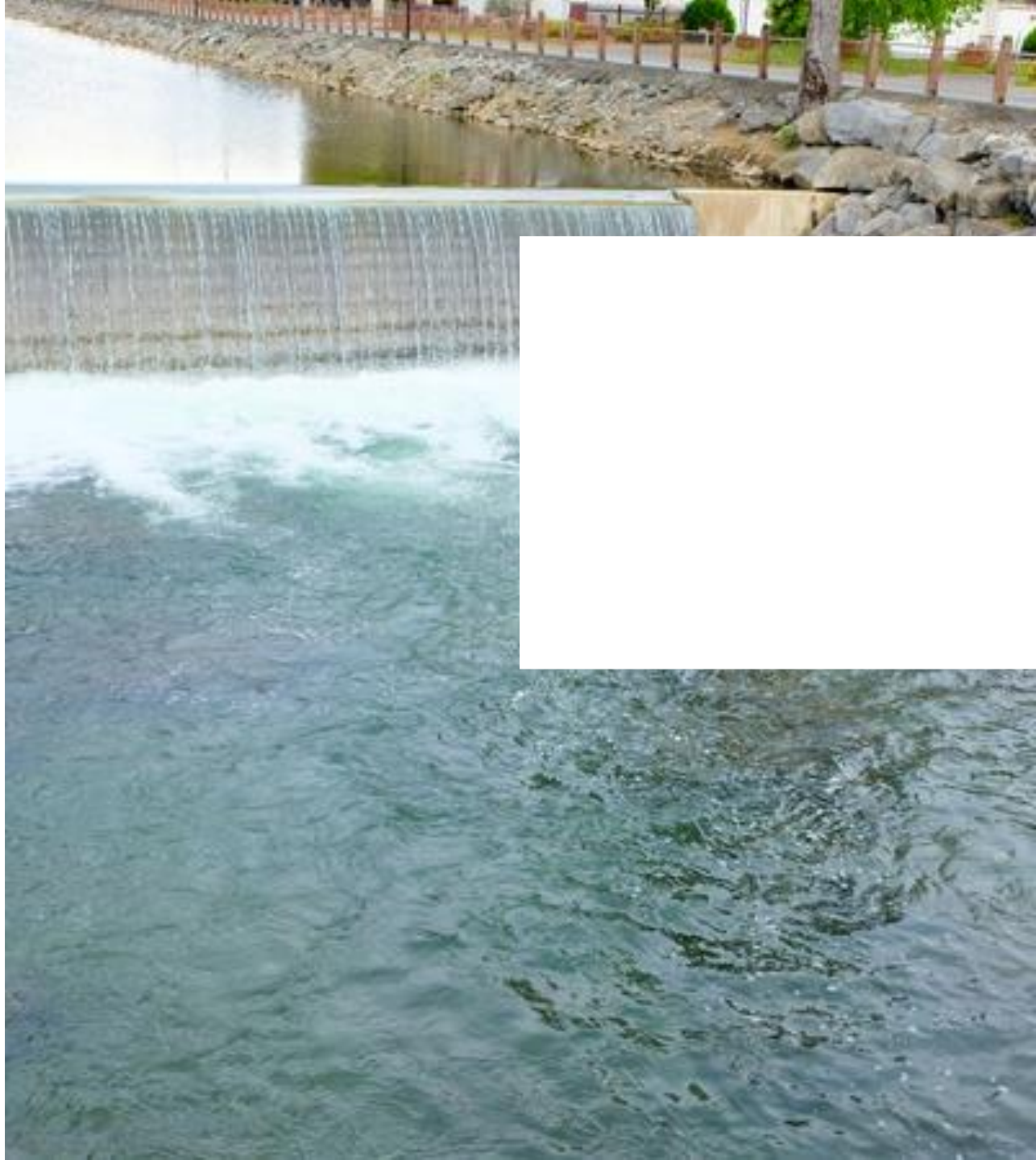
Bad debt collection in Utilities market

Analytics Assessment
2021

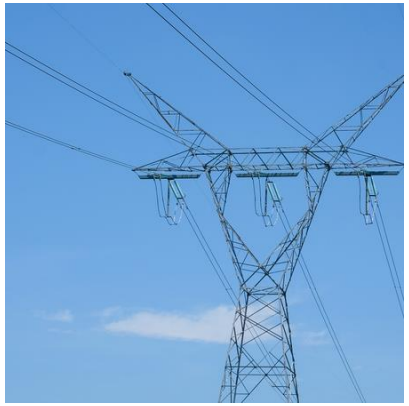
KEARNEY

I. Context
II. Guiding questions
III. Attachments

Context



Bad debt collection is a R$ MM operation for utilities companies, which can be significantly improved with Advanced Analytics



Context

- As in any pay-per-use service, Utilities companies have revenue losses due to defaulting customers
- Each year these companies spend millions of reais in their bad debt collection operation
- There are several collection actions that can be used to encourage customers to pay their bills, ranging from simple notices to fully disconnecting those customers from the network
- As these actions become more aggressive, they usually become more effective, but also more costly as well
- It's critical for these companies to select the appropriate collection action for each customer, so they can optimize this operation profitability
- Advanced Analytics models can significantly improve their selection methodology in order to achieve this goal

Exercise goals

- This exercise aims to evaluate **your rational** on the development of Advanced Analytics projects, as well as your **practical coding skills**
- To succeed in this exercise, **you must address every question** in the "Guiding questions" section
- The solution must be delivered in a python / R notebook format named **"[CANDIDATE_NAME].(notebook_extension)"**, alongside with a **.csv file with the predictions**
- Your solution should **include all your consideration, analyzes and charts** as part of this exercise
- The solution presented **will be discussed** in the technical interview
- The data available for this exercise is in a **file named utilities_disconnection.zip**. A description of this dataset columns is available in the appendix

Guiding questions

Guiding questions (1/2)

1

Initial Data Analysis

A. Exploratory Data Analysis

1. Read, clean and explore the received data. Explain each step and why.
2. What's the distribution for customer total debt variable? Draw a histogram. How should you group this variable?
3. Analyze the "collected" variable correlating it to separate variables. Plot this information.

B. Feature selection

1. Select the top 3-6 variables you consider that may be most important for root-cause analysis. Explain your considerations.
2. Are there any kind of preprocessing needed for these variables? Do that if so.

2

Model development and evaluation

A. Model selection

1. How do you select a Machine Learning model for this task? What kind of variable are you predicting? Select at least 3 different models.
2. What are the best hyperparameters for these models? Implement a hyperparameters tuning algorithm.
3. Create train-test splits for your data and implement a cross validation methodology for later evaluation.

B. Model evaluation

1. Which metrics can be used to evaluate the models' performance? Select at least 4 and explain their meaning briefly. Select the most appropriate one for model comparison.
2. Based on the cross validation in the training set, which model was more accurate?
3. Generate predictions for the test split for the best model. Join your predictions with the original info and save a file named **"[CANDIDATE_NAME]_PREDICTIONS.CSV".** How much is the success rate for this set?
4. Which variables were most important for this result? Calculate any kind of feature importance metric.

Guiding questions (2/2)

3

Business sense and technical background

A. Business sense

1. Describe how the client could use the model to improve its debt collection strategy.
2. Which additional data you think would most add predictive power to the model? Explain why.
3. How would you access this data?

B. Technical background

1. How do you avoid overfitting and underfitting in this problem?
2. Explain shortly how you would transfer your model to the client's IT Department. (Hint: DevOps)

Attachments

Exhibit 1
Available data columns description

Column	Section	Description
client_id	Id	Customer unique id
regional	Client info	Customer regional location id
site	Client info	Customer specific location id
client_class	Client info	Customer class id
client_subclass	Client info	Customer subclass id
connection_type	Connection info	Customer connection type (single-phase, biphasic, three-phase)
connection_status	Connection info	Customer commercial connection status at disconnection date (connected or disconnected)
installment_value	Debt history	Installment plan total value
installment_status	Debt history	Installment plan status (complete, current, not applicable)
total_debt	Debt history	Total debt at disconnection date
total_bills	Debt history	Quantity of bills overdue at disconnection date
disconnection_date	Disconnection info	Disconnection date
disconnection_requested	Disconnection info	Type of disconnection requested. There are 3 types of disconnection that can be executed in the field (in increasing order of complexity and aggressiveness: circuit break, pole, and branch). In some cases, it is possible to execute a disconnection remotely (remote)
contractor_id	Disconnection info	Contractor ID who performed the disconnection
disconnection_executed	Disconnection info	Type of disconnection executed (branch, pole, circuit break, remote)
collected	Target	Flag if debt was collected after the disconnection was executed

Thank you

Stay connected with Kearney

KEARNEY