

DMResources Limited - personalized micro credits

DMResources Limited is a recently created Fintech in Mexico which offers personalized micro credits to small and medium businesses. DMR has been able to capture their target market and achieved their desired market share in 20XX, however they have been experiencing a higher-than-expected default rate in their credit portfolio, this issue exposes DRM to regulatory liability and affects available capital due to constant changes in the required reserves.

The risk management unit of DRM has contacted Accenture to reengineer the default risk (churn) model to effectively comply with Mexican regulation and to reduce churn which directly translates into optimizing profit.

As Applied Intelligence analyst you are required to develop such a Data Asset to attend to your client's needs. Please take into consideration that any AI Data Asset should be built upon the pillars of **reproducibility, scalability and highest of qualities**.

The data you'll be working with consists of 3 tables provided by DMR. The data dictionary the sent is provided below:

- customerID: Customer ID
- gender: Whether the customer is a male or a female
- SeniorCitizen: Whether the customer is a senior citizen or not (1, 0)
- Partner: Whether the customer has a partner or not (Yes, No)
- Dependents: Whether the customer has dependents or not (Yes, No)
- Tenure: Number of months the customer has stayed with the company
- PhoneService: Whether the customer has a phone service or not (Yes, No)
- MultipleLines: Whether the customer has multiple lines or not (Yes, No, No phone service)
- InternetService: Customer's internet service provider (DSL, Fiber optic, No)
- OnlineSecurity: Whether the customer has online security or not (Yes, No, No internet service)
- OnlineBackup: Whether the customer has online backup or not (Yes, No, No internet service)
- DeviceProtection: Whether the customer has device protection or not (Yes, No, No internet service)
- TechSupport: Whether the customer has tech support or not (Yes, No, No internet service)
- StreamingTV: Whether the customer has streaming TV or not (Yes, No, No internet service)
- StreamingMovies: Whether the customer has streaming movies or not (Yes, No, No internet service)
- Contract: The contract term of the customer (Month-to-month, One year, Two year)
- PaperlessBilling: Whether the customer has paperless billing or not (Yes, No)
- PaymentMethod: The customer's payment method (Electronic check, Mailed check, Bank transfer (automatic), Credit card (automatic))
- MonthlyCharges: The amount charged to the customer monthly
- TotalCharges: The total amount charged to the customer
- Churn: Whether the customer churned or not (Yes or No)

Deliverable

Your project deliverable consists of two main components. Component 1 is a code project (R, Python, etc. Whichever you are more comfortable with) in which you process and model DMR data. Component 2 consists of a summary slide deck on which you report your findings to the clients.

Component 1 – Deepdive

Your code project must have, and will be evaluated as such, the following characteristics:

- A. Your code should be readable and understandable, any step in your workflow should be transparent
- B. Be completely reproducible in any computer (including random split seed)
- C. Contain at least a data folder and a code folder
- D. At least contain sections within your code for each step in your workflow
- E. A README file with an overall description of the problem statement, the technology/model used and a high-level description of the code functionality.
- F. All code should be documented, hard coded input logic explained, model hyperparameter tuning explained, cleaning steps and preprocessing logics etc.
- G. Organized methodology: Data Cleansing, transformation, testing/training, etc.

Additionally, any or of all the following are considered a “nice to have” element for your evaluation

- A. Compartmentalized code (one main/setup script running a script for each workflow step)
- B. Semantic versioning with Git/SVM
- C. Deliver using a private code repository in Github/BitBucket
- D. REnv/PyEnv
- E. Scalable model bidding (automated workflow to try multiple models and decide over a certain metric which is best).
- F. Saved summary tables and model configurations, subject to be loaded to view/reproduce results.
- G. Feature selection

Tasks

- A. Exploratory Data Analysis where you explore the relationship between churn and client characteristics. Grouped summary tables and plots are expected.
- B. Treat any pathologies on the data (for example: outliers, missing data, class imbalance, etc.)
- C. Propose and fit a classification model to predict customer churn (do not forget to split your data into train and test groups).

Component 2 – Presentation

What we are expecting to see:

- A. 3 – 5 minimal executive slide deck (Not including title slide or section slides)
- B. 1 slide, Intro/Your understanding of the problem, learn introduction of the study case
- C. 1 – 2 slides, EDA, Data findings derived from a clear data exploration process. (Relevant insight only)
- D. 1 – 2 slides, Model Choice & Results, Reliable machine learning model - you will be asked to explain the metrics used for the evaluation.
- E. Explanation of the results, why is your model the best model for the client?

Format

You will have maximum 10 minutes to explain your results (you will be given a warning at the 5- and 1-minute left time run and a hard stop at the 10 minutes mark), next you will have 5 minutes of Q&A.

Due date and delivery channel

Because DMR is in a rush due to an upcoming government audit, Accenture has committed to provide a complete data pipeline within a 72-hour timeframe. Your time will start running upon sending this email to your confirmed email address and will end 72 hours later.

Your work should be sent via email to your reporting Manager (your current recruiter). Any non-compliant elements of your deliverable will negatively affect your Accenture's performance evaluation (for example, dirty code, non-reproducibility, longer slide decks, non-required sections in deck, etc). Any work sent after the 72 hours timeframe won't be evaluated with no exceptions.