Nome da base de dados: Credit cart customers

Link para a base de dados: <https://www.kaggle.com/sakshigoyal7/credit-card-customers>

Fases do projeto:

|  |  |
| --- | --- |
| Contextualização | De onde a base de dados veio, ao que se refere, datas |
| Dicionário de Dados | Nome da coluna seguido de breve explicação sobre seu significado |
| Importação de Bibliotecas |  |
| Carregando base de dados e visualizando atributos | Os atributos são: shape, columns, head, tail |
| Análise inicial da base | A análise da base consiste, em um primeiro momento, em verificar a quantidade de instâncias preenchidas, instâncias nulas, os tipos de dados de cada atributo (info, isnull, visualização de colunas - unique, reconhecimento dos ‘Unkown’) |
| Análises estatísticas | A análise estatística permite a visualização da dispersão dos dados, análise de frequência através de histogramas, verificação de outlier com o uso do boxplot, e outras características como média, desvio padrão e percentis dos dados numéricos. |
| Tratando dados | O tratamento de dados é responsável por padronizar as instâncias de cada atributo. Nesta etapa, podemos preencher valores faltantes (nulos) quando necessário, alterar cada tipo de dado, realizar agrupamentos em categorias, dentre outros.  Aqui, aquelas colunas do Naïve Bayes podem ser retiradas talvez…  Se quiser, o impacto das transformações nas estatísticas, pode ser avaliado aqui |
| Perguntas para os dados | Aqui colocamos as perguntas e tentamos respondê-las com os dados já tratados |
| Conclusão | O que aprendemos e qual foi a importância do projeto (?) |

Colunas do arquivo:

|  |  |  |  |
| --- | --- | --- | --- |
| Nome da coluna | Explicação Kaggle | Explicação própria | Tipo |
| CLIENTNUM | Client number. Unique identifier for the customer holding the account | Identificador único para cada cliente da base de dados | Ordenar valores para checar |
| Attrition\_Flag | Internal event (customer activity) variable - if the account is closed then 1 else 0 | Se o cliente está ou não com uma conta ativa (se 1, a conta foi fechada, senão 0) | Categórica |
| Customer\_Age | Demographic variable - Customer's Age in Years | Idade do cliente | Discreta |
| Gender | Demographic variable - M=Male, F=Female | Sexo do cliente | Nominal |
| Dependent\_count | Demographic variable - Number of dependents | Número de dependentes do cliente | Discreta |
| Education\_Level | Demographic variable - Educational Qualification of the account holder (example: high school, college graduate, etc.) | Nível educacional do cliente | Ordinal |
| Marital\_Status | Demographic variable - Married, Single, Divorced, Unknown | Estado civil do cliente | Nominal |
| Income\_Category | Demographic variable - Annual Income Category of the account holder (< $40K, $40K - 60K, $60K - $80K, $80K-$120K, > $120K, Unknown) | Categoria de renda na qual o cliente se encontra (< $40K, $40K - 60K, $60K - $80K, $80K-$120K, > $120K, Unknown) | Intervalos |
| Card\_Category | Product Variable - Type of Card (Blue, Silver, Gold, Platinum) | Tipo de cartão do cliente (Blue, Silver, Gold, Platinum) | Ordinal |
| Months\_on\_book | Period of relationship with bank | Período de relacionamento do cliente com o banco (em meses) | Discreta |
| Total\_Relationship\_Count | Total no. of products held by the customer | Total de produtos do banco que o cliente possui | Discreta |
| Months\_Inactive\_12\_mon | No. of months inactive in the last 12 months | Número de meses que o cliente permaneceu inativo | Discreta |
| Contacts\_Count\_12\_mon | No. of Contacts in the last 12 months | Número de contatos nos últimos 12 meses | Discreta |
| Credit\_Limit | Credit Limit on the Credit Card | Limite do cartão de crédito | Contínua |
| Total\_Revolving\_Bal | Total Revolving Balance on the Credit Card | Média da quantidade gasta no cartão por mês | Contínua |
| Avg\_Open\_To\_Buy | Open to Buy Credit Line (Average of last 12 months) | Média de limite comprado nos últimos doze meses (reajuste do limite) | Contínua |
| Total\_Amt\_Chng\_Q4\_Q1 | Change in Transaction Amount (Q4 over Q1) | Comparação entre o uso do cartão nos 4 últimos meses do ano com os 4 primeiros meses do ano (em quantia) | Contínua |
| Total\_Trans\_Amt | Total Transaction Amount (Last 12 months) | Total de transferências nos últimos 12 meses (em quantia) | Contínua |
| Total\_Trans\_Ct | Total Transaction Count (Last 12 months) | Total de transferências nos últimos 12 meses (em número de transações) | Discreta |
| Total\_Ct\_Chng\_Q4\_Q1 | Change in Transaction Count (Q4 over Q1) | Comparação entre o uso do cartão nos 4 últimos meses do ano com os 4 primeiros meses do ano (em número de transações) | Contínua |
| Avg\_Utilization\_Ratio | Average Card Utilization Ratio | Razão média de uso do cartão (Rotativo/Limite de crédito) | Contínua |

\* Acrescentar tipo de variável:

Categórica

Quantitativa ordinal

Quantitativa nominal

\* Para as colunas categóricas, colocar o unique().