

# ***Unsupervised Learning*** **aplicado no Banco Beta**

**Eduardo Germano<sup>1</sup>**

<sup>1</sup>Especialização em Data Science  
UniRitter – Brazil

`eduar.germano@gmail.com`

O Banco Beta é uma sólida instituição financeira no cenário nacional. A instituição está passando por um processo de expansão e necessita adotar estratégias que apoiem este crescimento. Com a expansão da carteira de clientes, torna-se necessário criar ações personalizadas para cada tipo de perfil de cliente do banco. Deste modo, Maria tem a oportunidade de utilizar aprendizagem de máquina não-supervisionada [Barlow 1989] para segmentar a carteira de clientes do Banco Beta.

Para o sucesso de sua tarefa, Maria deverá, primeiramente, agrupar as informações dos clientes em uma base e limpar os dados de forma que a mesma possa prosseguir com sua tarefa. Em seguida, Maria poderá utilizar a técnica de *PCA* (*Principal Component Analysis*) [Maćkiewicz and Ratajczak 1993] para identificar as características mais relevantes nos dados dos clientes e reduzir a dimensionalidade do modelo de conhecimento em questão. Tal técnica tem como objetivo reduzir e otimizar o modelo de conhecimento.

Após a preparação dos dados, Maria pode aplicar o algoritmo *K-Means* [Hamerly and Elkan 2004] para gerar a segmentação da carteira de clientes em  $N$  grupos. Tal segmentação também é conhecida como clusterização, que é o particionamento dos dados em grupos com instâncias de características similares. A utilização do *K-Means* em nosso *case* é mais adequada, pois estamos lidando com dados não-rotulados e desejamos que o algoritmo extraia o conhecimento em nosso modelo. Além disso, o *K-Means* é muito utilizado para tarefas de clusterização [Burkardt 2009].

Um ponto importante é que um especialista do negócio deverá analisar uma amostra de cada *cluster*, com o objetivo de entender a característica de cada segmentação criada para a carteira. Com a carteira devidamente segmentada, o Banco Beta poderá fazer comunicações específicas, concessões de crédito personalizadas, e testes A-B para entender o comportamento dos clientes de cada segmentação. Os resultados destas ações deverão ser bem analisadas, de modo que sejam identificadas as melhores estratégias para retenção dos clientes de cada grupo.

Por fim, após compreendido a característica de cada *cluster*, Maria também poderá treinar um algoritmo de aprendizado supervisionado para classificar novos clientes em um dos segmentos da carteira do banco. Deste modo, o Banco Beta estará preparado para enfrentar o intenso processo de expansão de sua carteira de clientes, e se manter como uma instituição financeira sólida no cenário nacional.

## **Referências**

Barlow, H. B. (1989). Unsupervised learning. *Neural computation*, 1(3):295–311.

- Burkardt, J. (2009). K-means clustering. *Virginia Tech, Advanced Research Computing, Interdisciplinary Center for Applied Mathematics*.
- Hamerly, G. and Elkan, C. (2004). Learning the k in k-means. *Advances in neural information processing systems*, 16:281–288.
- Maćkiewicz, A. and Ratajczak, W. (1993). Principal components analysis (pca). *Computers & Geosciences*, 19(3):303–342.