

Alunos: Victor Silveira Portelinha, Joel Sepulveda Martins, Lucas Azevedo Dias, Lucas Gabriel Nunes Geremias

- Determine se é um cenário de Big Data ou não. E, caso seja, justifique sua resposta considerando os 5 Vs.
- Determine se os dados armazenados por esta empresa são estruturados, semiestruturados e/ou não-estruturados.
- Determine se os dados são provenientes de fontes internas ou externas da empresa.

1. Análise, em tempo real, do número de reclamações sendo geradas e atendidas pelos diferentes meios, isto é , telefone, web e mobile.

- 1.1 R: é Big Data pois a vários dados sendo gerados ao mesmo tempo, com vários tipos, em grande velocidade, com necessidade de veracidade pois as reclamações podem não servir ou serem falsas.
- 1.2 R: Não estruturados pois mensagens, e gravações no telefone não possuem nenhuma estrutura definida de informações
- 1.3: R: Os dados são provenientes de fontes internas, pois os dados são salvos num sistema interno.

2. Identificação usando modelos probabilísticos, também em tempo real, de potenciais fraudes das transações realizadas em cartão de crédito;

R: No tocante à análise de uma problemática advinda da área de Big Data, é notável o tratamento de uma massiva quantidade de dados, de diversas fontes, onde consequentemente, deverá ser feito um tratamento e processamento destes, retornando através de um algoritmo de probabilística, uma síntese estatística do problema então solicitado. Em suma, os V's encontrados seriam o de volume, como citado acima pela grande quantidade de dados, também o de variedade, visto que o serão dados gerados e comparados de diversas fontes em ligação com este esta a velocidade, que de forma intuitiva é notável a importância do processamento em tempo real, para a detecção de uma possível fraude, além destas existe a veracidade, que garante uma análise próxima a realidade para que não ocorra decisões errôneas

semi estruturados, devido a tipos de dados muito diferentes entre si

Internas: Histórico de transações

Externas: Dados de transações, históricos

3. Cruzamento de informações para identificar potenciais empresas lavadoras de dinheiro;

- 3.1 R: É um cenário de Big Data, por causa do volume de dados sensíveis que deverão ser cruzados em um curto intervalo de tempo de forma confiável.
- 3.2 R: Serão usados dados estruturados.
- 3.3 R: Os dados são de procedência interna.

4. Determinar, em tempo real, usando modelos estatísticos, a probabilidade de que um cliente requisitante de crédito, entre em inadimplência;

4.1 Sim, é Big Data. Há um grande volume de dados a serem analisados, o que gera problemas de volume e velocidade. Além de os dados a serem analisados são em grande parte semi-estruturados, gerando uma grande variedade em sua análise. Seria preciso criar sua própria solução/algoritmo para que esses dados possam ser analisados.

4.2 Os dados são semi-estruturados. O banco já possui diversas fontes de informações prévias de um cliente em específico, como seu nome, idade, padrões de compra dentro de sua conta no banco, seriam dados estruturados que poderiam ser facilmente encontrados dentro do SGBD do banco. Dados não estruturados seriam fornecidos por terceiros, como por exemplo, postagens em redes sociais ou cookies minerados em outros sites da web.

4.3 Os dados são providos tanto de fontes externas quanto internas.

5. Análise dos conteúdos gerados pelos clientes do banco em redes sociais, visando angariar informações sobre a satisfação ou insatisfação dos seus clientes e seus motivos.

- 5.1 R: Acaba sendo um cenário de Big Data, justamente por causa do volume de dados que serão analisados em um curto intervalo de tempo tendo foco na confiabilidade desses dados.
- 5.2 R: São dados semi-estruturados e não-estruturados.
- 5.3 R: São dados provenientes de fontes externas.