

Descrição

O caos decorrente do corona vírus já trouxe uma série de mudanças em relação ao mercado financeiro, principalmente no que tange aos canais de relacionamento entre as instituições financeiras e seus clientes. De um lado, temos o aumento de transações em canais digitais, o que traz mais agilidade e eficiência no relacionamento com os clientes, bem como redução nos custos operacionais quase eliminando a necessidade de uma agência física. Do outro lado, temos criminosos aproveitando o cenário de pandemia para tentar encontrar lacunas e cometer crimes financeiros como ofertas/transações fraudulentas e lavagem de dinheiro.

Desafio: Descreva como você utilizaria Ciência de Dados para evitar golpes e mitigar o risco de crimes financeiros nesse novo contexto. Reflita sobre quais seriam os atos ilícitos possíveis. Discorra sobre quais seriam as principais etapas do projeto e quais seriam os participantes. A seguir, proponha quais grupos de dados seriam relevantes para endereçar esse tema. Cite também ferramentas e técnicas que você utilizaria no processo de desenvolvimento de uma solução analítica. Finalmente, explique como você integraria a solução desenvolvida de forma produtiva nos canais de atendimento.

RESPOSTA

Descreva como você utilizaria Ciência de Dados para evitar golpes e mitigar o risco de crimes financeiros nesse novo contexto:

Há dois componentes principais no combate às fraudes financeiras que é a detecção fraudulenta, ou seja, o ato de pegar o larapio no pulo, no momento que está acontecendo o evento fraudulento (delito), sempre buscando padrões atípicos, aquilo que não é normal para o titular da conta. E a prevenção, essa é um pouco mais complicada porque busca analisar e prevenir os eventos fraudulentos, antes que ocorram. Desse modo ao analisar o histórico, faz o treinamento dos algoritmos computacionais através do treinamento supervisionado para que possa fazer ou tentar fazer a previsão, lá no futuro. O que vai acontecer lá, mediante os dados que se tem hoje. O cliente vai ser capaz ou não de honrar seus compromissos? Essa é uma das perguntas a ser respondida. Mas como saber? Como definir ou separar os bons dos maus pagadores? Essa provavelmente seria uma outra pergunta que em algum momento teria que ser respondida. Bem, prever um ato futuro não é uma tarefa fácil como em uma máquina que se pode prever quando uma determinada peça vai ou não quebrar por meio da fadiga, e antes que o motor possa fundir acontece a manutenção preventiva e retira aquela peça defeituosa para evitar que o motor quebre e o prejuízo seja ainda maior. Então, como prever os atos de uma pessoa? Como separar um ato lícito do ilícito? Bom, por mais que Direito nos fale e mostre que todos somos iguais, cada pessoa é diferente do outro, o que nos define o quão somos diferentes não são os nossos padrões comportamentais, mas sim os dados no geral. É como se, cada ser humano fosse um mundo à parte ou que fosse construído por um algoritmo diferente, o que torna o ser humano bom ou ruim é justamente a sua trajetória, como em um rio tortuoso buscando o caminho para o mar. O que define o ser em um determinado momento são as suas experiências passadas, os dados que foram gerados. Então uma forma de prever como aquele indivíduo vai se comportar mediante uma determinada situação pode estar relacionada com a sua experiência passada, os dados já gerados anteriormente. E nos dias de hoje por meio das redes sociais, informações do Serasa, testes comportamentais e uma infinidade de outras coisas, todas elas ou uma parte pode dizer muito como um determinado ser humano é, então para

prever uma ação futura se faria necessário analisar aquilo que já tem disponível, treinar os algoritmos, construir os modelos para que possam auxiliar na tomada de decisão de novos clientes, de empréstimos ou outros produtos por meio dos seus *insights*.

Refleta sobre quais seriam os atos ilícitos possíveis:

Vou citar 3 sendo eles o Phishing, Vishing e Smishing que são os golpes mais comuns para roubar informações dos clientes. **Phishing** é um golpe aplicado geralmente por mensagens de e-mail, nas redes sociais e sites falsos. O criminoso tenta atrair a vítima com uma informação falsa, fazendo com ela mesma entregue seus dados. A pessoa que clica em um anúncio desse, geralmente cai em um site que parece muito com o da loja real, o perigo mora quando ela se interessa por uma mercadoria e preenche um cadastro de compra com seus dados. Uma maneira de tentar evitar isso seria trabalhar com informação junto ao seu cliente investindo em *marketing* para avisar sobre os perigos decorrentes a essa prática. Outra maneira poderia se vincular o domínio do banco as contas do Google, tipo (nome_do_cliente@cliente.bradesco.br) desse modo poderia limitar ao máximo o envio de mensagem com o nome do banco para o cliente. Talvez ficaria caro demais que não compensaria. Ou vincular uma conta já existe de um cliente ao programa de *spam* do banco para que ele não receba tais mensagens como na imagem 1 e 2 onde os spams são filtrados no servidor POP/IMAP do banco.

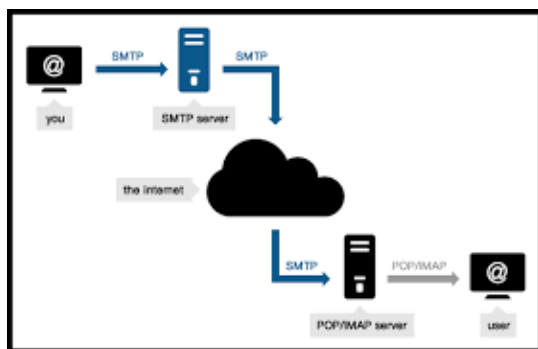


Figura 1 - recebimento de mensagens pelo provedor do cliente

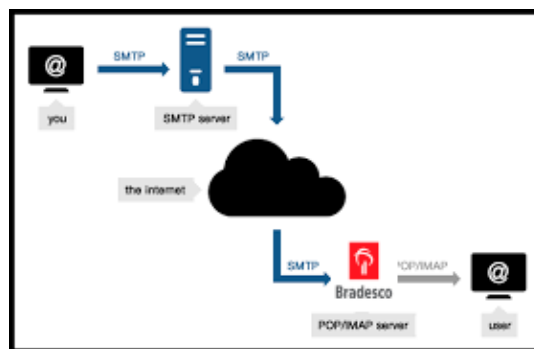


Figura 2 - Recebimento de mensagens pelo provedor do Banco

Vishing: É um golpe aplicado por áudio, nesse caso a pessoa recebe uma ligação para "confirmar" informações que provavelmente o criminoso não tem, convencendo a vítima a passar tais informações a ele, nesse caso não sei como impedir tal ligação, teria que ver com o pessoal da telefonia e impedir que determinado número fosse usado nesses golpes.



Smishing: Esse golpe vai pôr SMS sempre usando de chamariz seja com uma promoção imperdível ou oferta de crédito, fornecendo um *link* ou pedindo informações. Outra medida preventiva poderia adotar a interceptação dos SMS de modo que houvesse uma filtragem desses tipos de informação para o cliente final. Não sei se isso é legal ou moral. Mas poderia ser uma opção como demonstrado na imagem ao lado que mostra os SMS sendo filtrados ao passar pelo crivo do banco, desse modo garantiria que só mensagens que não fossem SPAM passariam.

Em todas as 3 práticas ilícitas citadas o objetivo é o mesmo roubo de senhas e informações do usuário. Mas se não tivesse essa senha ou se ela fosse substituída por reconhecimento facial ou

vocal, aplicado também ao registro da digital no momento que se usa o aplicativo para confirmar uma compra. A bastante tempo já tem telefones com reconhecimento da digital, desse modo o criminoso teria que quebrar o sistema do banco e não ir para o lado mais fraco que é o cliente onde está sujeito a informações falsas e ser enganados por criminosos. Ao fazer uma compra o aplicativo gravaria por alguns instantes o usuário falando uma frase, pegaria não só os sotaques mas também o padrão da fala do usuário, além do reconhecimento facial e da digital. Tudo para confirmar uma compra de um café? Hmm! O cliente não acharia muito? Concordo, teria que medir todos os pros e contra. Fazer degraus do tipo, usar o cartão para compras diárias, corriqueiras do dia a dia, ter um limite para isso e aquilo. E em relação a outras coisas usar o telefone com os protocolos já existentes de segurança aliados a outros para compras mais caras. Há tantas possibilidades para pensar e projeções a fazer que algumas práticas seriam inviáveis, mas o que fazer? Deixar os criminosos praticarem seus atos e ficar sempre na defensiva ou contra atacar usando os meios que já existem limitando suas ações e os frustrando. Com certeza eles se reinventariam para burlar a segurança. É preciso ter em mente que a medida que o mundo vai se modernizando seja com novas técnicas ou equipamentos sendo desenvolvidos para alcançar um determinado fim, a fraude caminha junto, ambos são sistemas dinâmicos evolui com tempo. Os criminosos são pessoas inteligentes isso não resta dúvida, penso que, um dia não haverá sistemas operacionais dentro de dispositivos como há nos smartphones atuais, mas sim, um hardware que se conecte a um sistema online e que tenha tudo que o usuário precise, hoje já existem Dockers que são *softwares* contêiner, talvez no futuro os sistemas operacionais (Android, IOS, Xamarim) esteja na nuvem e cada aparelho de usuário seja capaz de acessar seu sistema na nuvem e as informações poderão ser filtradas e saber de onde está vindo tais dados, podendo localizar a fonte com mais precisão. Hoje as ISO de vários sistemas são baixadas aos montes da internet e instaladas nos celulares, algumas podem até estarem corrompidas por programas maliciosos que ficam escondidos esperando as pessoas baixarem seus programas dos bancos e digitarem suas informações.

Discorra sobre quais seriam as principais etapas do projeto e quais seriam os participantes. A seguir, proponha quais grupos de dados seriam relevantes para endereçar esse tema.

As etapas consistem em transformar dados em *insights* como empresas de grande porte coletam muita informação é preciso se trabalhar com Big Data pois seu uso vem se tornando cada vez mais um diferencial competitivo para as empresas pois permite a transformação de dados em conhecimento (*insights* relevantes), muito importante para as tomadas de decisão. A seguir foi listado as etapas para se conseguir os *insights*

Coleta de Dados: Como os bancos estão sempre preocupados em oferecer aos seus clientes produtos e serviços, o setor de *marketing* pode se beneficiar com informações relacionadas ao comportamento do consumidor. Mesmo que a empresa adote a ideia de armazenar diferentes dados para possíveis aplicações futuras, é essencial garantir que eles cubram ao menos os critérios relevantes para o objetivo principal, então desse modo o *marketing* para varejo poderia coletar os seguintes dados:

- 1) Sistema operacional do dispositivo (isso diz muito sobre o cliente) e o tipo do dispositivo se é computador ou *smartphone*.
- 2) Cliques em anúncios
- 3) *Login* do cliente no site, caso haja.
- 4) Endereço de IP
- 5) Seu *e-mail*

- 6) Histórico de buscas e de compras
- 7) Redes sociais
- 8) Endereços físicos
- 9) Número de telefone
- 10) Informações de perfil
- 11) Consulta ao SPC/Serasa além de outros

Limpeza de Dados: A limpeza de dados é um pré-processamento dos dados, é aqui que se identifica anomalias que por ventura possa comprometer a análise. O que se faz ou não aqui reflete na visualização dos dados, então essa é uma etapa crucial. Porque aqui são retirados os valores nulos, as inconsistências, os valores duplicados além de outros, onde essa limpeza permite que a base de dados seja enriquecida sugerindo novos parâmetros para a coleta de dados como:

- 12) A localização geográfica
- 13) Registros criminais
- 14) Hábitos de compra
- 15) Histórico médico
- 16) Histórico de crédito.

Mineração de Dados: Essa etapa tem como missão a identificação de padrões onde são utilizados métodos de Inteligência Artificial (gosto do termo: inteligência Computacional), *Machine Learning*, estatística além de outros. O Big Data não olha dados comportamentais, mas sim para os dados em geral fornecendo *insights* relevantes que um olho humano não veria. Afinal quem pensaria em oferecer vinhos chilenos para uma pessoa que compra uma determinada marca de xampu? Parece ilógico isso, mas para o Big Data não importa se a relação lógica está ou não explícita, o fato é que ter essa informação pode beneficiar o varejista.

Análise de Conteúdo: Essa etapa consiste em analisar de forma eficiente o conteúdo das informações que foram coletadas, filtradas e validadas é o que vai realmente gerar *insights* benéficos para o empreendimento. Para diferentes objetivos essa etapa pode ter diferentes métodos. Por exemplo, para entender o cenário econômico no qual a empresa atua, pode-se fazer uma análise descritiva focada no presente ao fornecer dados em tempo real sobre o desempenho da empresa. Por outro lado a preditiva visa identificar padrões para prever possíveis cenários futuros e dar base para as decisões a serem tomadas. Uma outra análise chamada de prescritiva mostra ao gestor possíveis resultados de uma ação tomada em um determinado cenário, auxiliando no ajuste da estratégia para que os objetivos sejam alcançados. Por último não menos importante que os demais, tem-se a análise diagnóstica que ajuda a elucidar cadeias de acontecimentos relacionados à ocorrência de uma falha em um processo.

Visualização de informações: Nessa etapa, o desafio é tornar o acesso ainda mais fácil a todas as pessoas envolvidas na operação gerando uma visualização intuitiva dos dados com gráficos, infográficos, planilhas e tabelas. Contudo é fundamental conhecer as pessoas que terão acesso aos dados. Além disso, é preciso avaliar o que exatamente elas buscarão naquele documento. Uma mesma análise de Big Data pode resultar em diferentes visualizações, cada uma visando a facilitar a interpretação de um determinado “público-alvo”, ou seja, gestores de diferentes departamentos.

Integração de dados: Essa etapa significa dizer que a integração dos dados será utilizada de forma sistemática e com inteligência. Isso significa que não basta executar análises pontuais e independentes. Veja bem, hoje em dia, não faz sentido encarar uma empresa como uma gama

de departamentos com funcionalidades distintas. É preciso entender a importância de contar com setores alinhados em seus objetivos e que, portanto, podem ter certas funções integradas. Sendo assim, as etapas de Big Data devem gerar a colaboração entre todas as partes. De acordo com as características do negócio e os objetivos que estão sendo perseguidos as etapas de Big Data podem variar. Contudo, pontos como o estabelecimento de metas, o garimpo, a validação das informações e a integração de dados são elementares em qualquer abordagem. Seguindo esses passos, abre-se caminho para a geração de *insights* que possibilitam decisões mais efetivas.

Cite também ferramentas e técnicas que você utilizaria no processo de desenvolvimento de uma solução analítica.

O grande volume de informação e plataformas de estudo na análise de Big Data pode deixar o processo de escolha confusa, tornando um grande desafio saber qual a técnica e ferramenta ideal para cada objetivo ou contexto. Eu não sei se o Bradesco usa servidores próprios para processar seus dados ou se usa a nuvem, mas atualmente três das maiores empresas estão lutando para hospedar aplicativos do futuro sendo elas a Amazon, Google e Microsoft. Uma matéria que saiu no canaltech indica que o Azure é superior ao Google Cloud e a Amazon devido a utilização de *backbone* para o transporte de dados, enquanto que o Google e a Amazon utilizam a internet. Acredito que o processo para tratamento dos dados e obtenção dos *insights* deva ser o mesmo para qualquer uma das plataformas, o que muda é qual serviço será usado. O Google tem o BigQuery que separa armazenamento e computação para ativar o escalonamento elástico e fornecer *insights* rápidos baseados em petabytes de dados, e para analisar os dados pode-se usar ferramentas famosas de BI como Tableau, Looker, Qlik ou Planilhas Google. De acordo com o (help.tableau.com) também tem jeito de instalar o Tableau no Azure por meio de máquinas virtuais. Poderia se usar também o Apache Hadoop, Spark, Skytree Server, Weka/Pentaho e por aí vai.

Em relação as técnicas de análise de Big Data, tem-se a Mineração de Dados, análise de redes sociais, *web mining*, *machine learning* e métodos de otimização. Falo um pouco delas a seguir:

Mineração de dados: é o processo de descobrir padrões e relações úteis, em grandes volumes de dados. Podendo combinar ferramentas de estatística e inteligência artificial (como redes neurais e *machine learning*) com gerenciamento de banco de dados para analisar grandes conjuntos de dados. As ferramentas usadas para isso poderia ser o excel, o Knime, Weka/Pentaho, Rapid-I RapidMiner, sempre uso ferramentas open-source como o software R, python e suas bibliotecas. O software R é tão bom que até o Power BI desktop tem ele para analisar dados.

Análise de Rede Social: Essa técnica é utilizada para ver as relações sociais no ambiente das Redes Sociais podendo ser quantitativa geralmente focando em bases de dados preexistentes, e qualitativas sendo feita através de entrevistas, questionários ou mesmo da observação do pesquisador de um determinado ambiente ou grupo. As ferramentas para coleta de dados são denominadas de *crawlers*. Para Twitter, Facebook, Youtube e Flickr tem o NodeXL que não precisa de servidor, tem também o NetVizz, o Gephi, NetDraw e outras.

Web mining (Mineração da Web): é uma técnica empregada para descobrir padrões em grandes repositórios da web

Machine Learning: As técnicas de *Machine Learning* permitem que os usuários façam previsões a partir de grandes conjuntos de dados, desenvolvendo técnicas de aprendizagem eficientes

(algoritmos), dados ricos e grandes ambientes de computação. Assim, essa técnica tem um grande potencial e é uma parte essencial da análise de big data.

Métodos de Otimização: Os métodos de otimização são usados em campos multidisciplinares sendo utilizados para resolver problemas quantificáveis. Podendo usar o Matlab, o Python, o Software R além de outros.

Finalmente, explique como você integraria a solução desenvolvida de forma produtiva nos canais de atendimento.

Hoje em dia não faz sentido ter uma empresa que tenha departamentos com funcionalidades distintas, é preciso que toda a equipe tenha certas funções integradas, deste modo as etapas de Big Data devem gerar a colaboração entre todas as partes. Isso de acordo com as características do negócio e os objetivos que estão sendo perseguidos. Nesse passo, pontos como o estabelecimento de metas, o garimpo, a validação das informações e a integração de dados são elementares em qualquer abordagem. Seguindo esses passos, abre-se caminho para a geração de insights que possibilitam decisões mais precisas.