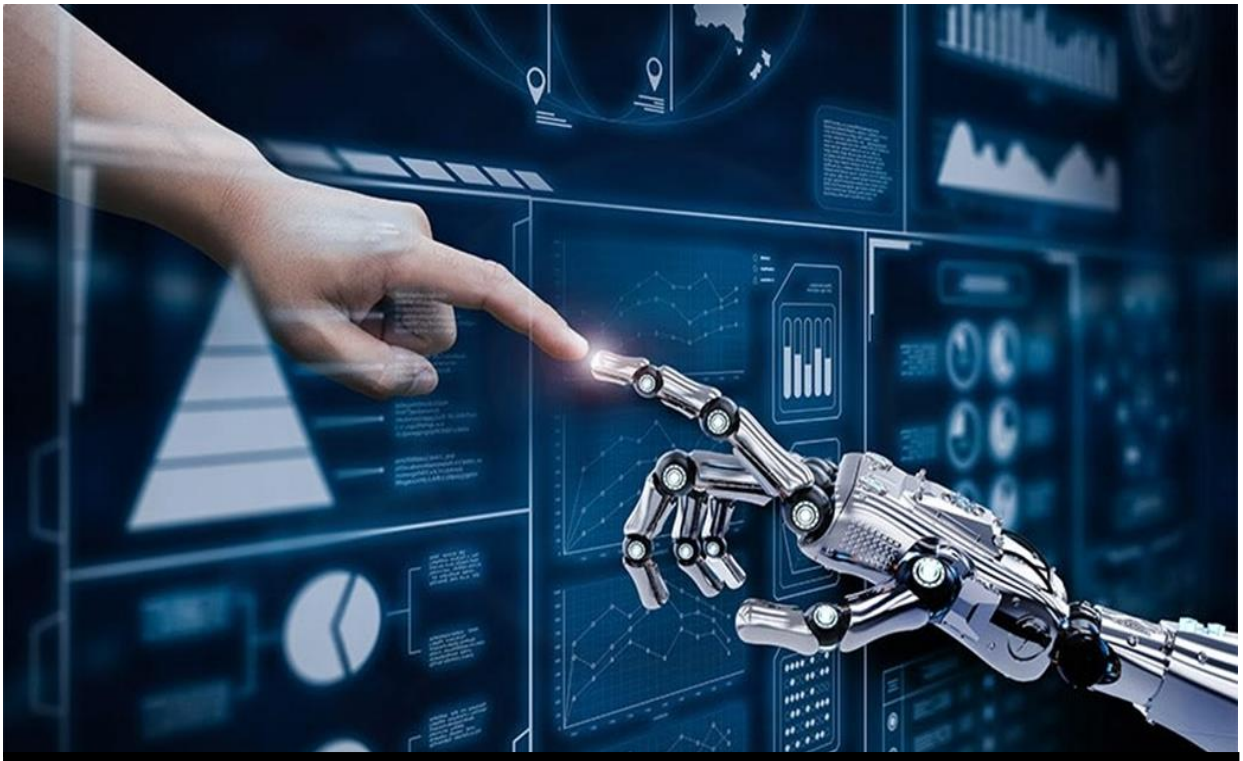
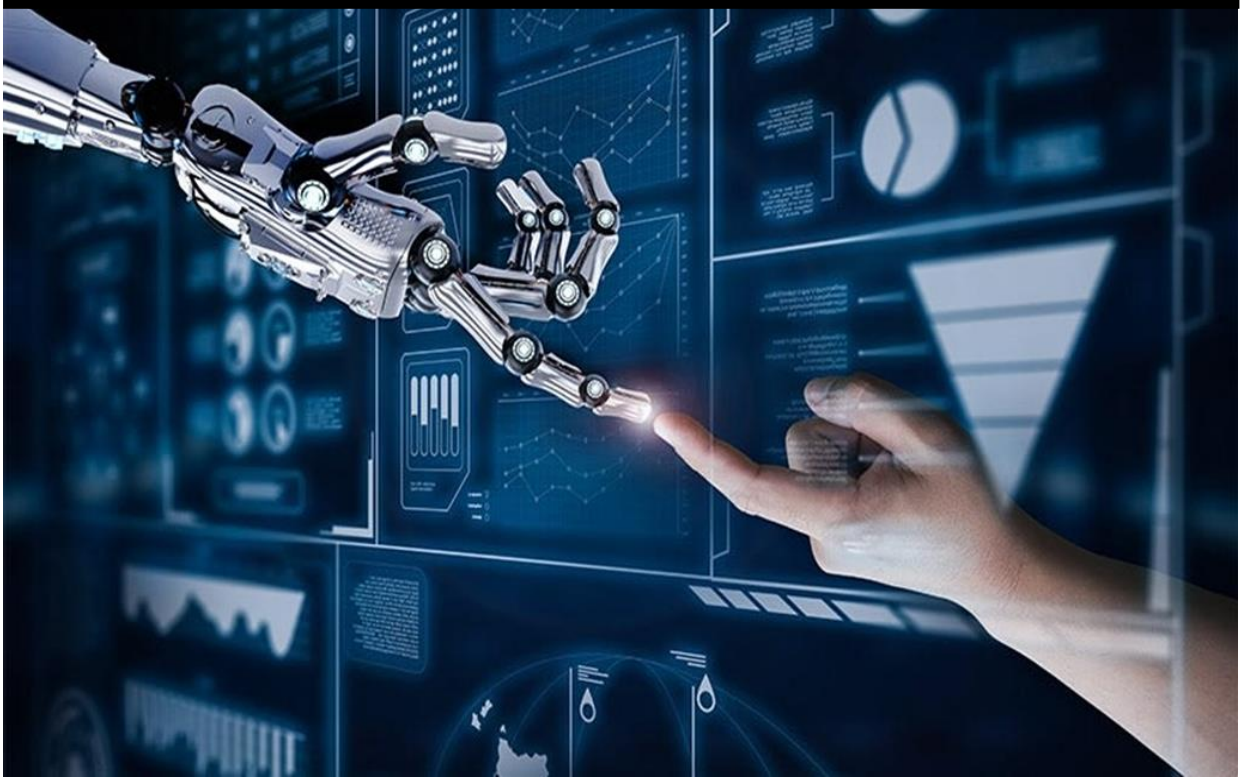


Mentoria DSA- 2021



1. APRESENTAÇAO EQUIPE DE MENTORIA

- **Anne Francine Martins (Mentora)**

Doutora em engenharia industrial pela Arts et Métiers (Paris/ França), mestre em pesquisa operacional e economista pela Universidade Federal do Espírito Santo. Atuo como cientista de dados e product owner data na 42 C e também sou professora de mestrado em Big Data / Data Analytics na Estiam (Paris /França) .

✓ PROJETOS: Mentoria Data Science Academy 2021
https://blog.dsacademy.com.br/programa-de-mentoria-entre-os-alunos-dsa-temporada-2021/

✓ Coordenadora Projeto de Discurso de ódio / Racismo nas Redes Sociais junto com PUC/SP

LinkedIn: https://www.linkedin.com/in/anne-francine-martins/
GitHub: https://github.com/annefrancine

- **Luiz Henrique de Oliveira Bueno (Mentor)**

Graduação em Logística (UNC); Pós Graduação em Engenharia e Gestão da Produção (UNICESUMAR); Data Scientist (DSA) ; MBA- Ciência de Dados (PUC-RIO) ; Engenharia de Software para Ciência de Dados (PUC-RIO) ; Analista de Machine Learning (IGTI), Scrum Master Web Developer.

EXPERIÊNCIAS: Technology Instructor Data Science IMPACTA Tecnologia ; Consultoria e Pesquisa Data Science ; Membro Associado I2AI Associação Internacional de Inteligência Artificial ; Membro do Comitê de Machine Learning I2AI.

PROJETOS: Ebook Desmistificando Machine Learning https://conteudo.i2ai.org/desmistificando-o-machine-learning

Mentoria Data Science Academy 2021
https://blog.dsacademy.com.br/programa-de-mentoria-entre-os-alunos-dsa-temporada-2021/

LinkedIn: www.linkedin.com/in/luiz-henrique-sc

- **Natalia Faraj Murad (mentorada)**

Graduada em Ciências Biológicas (UFLA); Mestrado em Genética e Biologia Molecular (UNICAMP); Doutorado em Genética e Biologia Molecular (Unicamp). Atualmente é pesquisadora de Pós-Doutorado na área de Bioinformática e Data Science na Unicamp.

PROJETO 1: Modelo machine learning - Previsão churn https://github.com/dsamentoria/analise_churn

PROJETO 2: Previsão bitcoin com Séries Temporais
https://github.com/dsamentoria/bitcoin

Linkedin: https://www.linkedin.com/in/natmurad

GitHub: https://github.com/natmurad

- **Rafael Henrique Gallo (mentorado)**

Estudante engenharia da computação - 5º Semestre faculdade IMPACTA.

EXPERIÊNCIAS: Estagiário data Science Gentrop Cloud Brasil 2021 - Atualmente atua com criação de modelos machine learning, análise de dados com BigQuery no Google Cloud. Atuando em projetos de marketing cloud com machine learning, análise de dados.

PROJETO: Modelo machine learning - Previsão churn
https://github.com/dsamentoria/analise_churn/blob/main/Rafael%20Gallo%20-%20Modelo%20machine%20learning/Jupyter/Model%20M.L%20-%20Previs%C3%A3o%20Churn.ipynb

Modelo machine learning - Previsão bitcoin com ARIMA SARIMA
https://github.com/dsamentoria/bitcoin/blob/main/Gallo%20Raf/Projeto%2002%20-%20Modelo%20ARIMA%20SARIMA%20Bitcoin1%20(1).ipynb

LinkedIn: https://www.linkedin.com/in/rafael-gallo-986a73150/

GitHub: https://github.com/RafaelGallo

Website: https://rafaelgallo.github.io/webportfolio/

2. O QUE É MENTORIA?

A tutoria, também chamada de mentoring, é um método muito utilizado para efetivar uma interação pedagógica. Os tutores acompanham e comunicam-se com seus alunos de forma sistemática, planejando, dentre outras coisas, o seu desenvolvimento e avaliando a eficiência de suas orientações de modo a resolver problemas que possam ocorrer durante o processo.

O tutor pode ser, ele próprio, ainda um estudante. Este fato tem a vantagem de propiciar um contato menos formal junto do aluno tutorado de forma a que a mensagem transmitida pelo tutor seja mais rapidamente compreendida e assimilada o que facilita o acesso ao conhecimento, e que numa relação demasiado formal poderá ser dificultada ou mesmo impedida.

A aprendizagem tutorial exige estrutura predeterminada e predefinida. Este tipo de aprendizagem tem suas raízes no cognitivismo. O tutor guia o tutorado com auxílio de um fio condutor que atravessa uma grande parte das disciplinas.

O tutor conhece as necessidades e soluções, pelo fato de ter vivenciado semelhantes dificuldades e por conhecer formas de superá-las. Ele pode ser um grande amparo para o aluno em todo o momento em que o aluno tutorado estiver sobrecarregado, intervindo e auxiliando-o.

Esta estratégia de condução da aprendizagem agrada muitos alunos, porque sentem-na pouco restritiva, pouco limitadora, simplesmente porque acabam por aprender a dominar um bloco de disciplinas de maneira muito eficiente, e por ser-lhes, também, dada uma série de dicas sobre métodos de estudo e formas de apreensão das matérias.

3. O QUE É UM PROJETO?

Segundo a ISO 10006, um projeto é um processo único, que consiste num conjunto de atividades coordenadas e controladas com datas de início e fim, realizadas com o objetivo de atingir um objetivo de acordo com requisitos específicos como restrições de tempo, custos e recursos.

Um projeto possui algumas características:

● Ser temporário é único: Um projeto é limitado no tempo e no espaço. Ele possui um início e um fim e ocorrem dentro de um contexto específico. Temporário não significa curto prazo. Alguns projetos podem durar vários anos, mas todos têm um fim pretendido quando os objetivos do projeto são alcançados

● Possuir início, meio e fim: Para gerenciar com sucesso um projeto, você deve dividi-lo em várias fases e colocar essas fases no tempo.

● Definir um budget: realizar um estudo preliminar de custos e benefícios ou receita esperada, estudo de riscos operacionais e financeiros e vários impactos, etc.

● Objetivos: Cada projeto deve definir objetivos claros formalizando a necessidade do cliente.

● Incerteza: Em qualquer projeto existe um elemento de incerteza ligado ao caráter unitário, mas também ao ambiente externo que pode ser difícil de apreender.

● Inovação: O projeto vai além da mera redução de custos, promovendo a inovação e consequentemente a criação de valor. Em outras palavras, a inovação leva à maximização da criação de valor. A essência de um projeto é a inovação.

● Um resultado: O produto. O projeto serve para atender a necessidades bem definidas e julgar seu sucesso.

3.1 GESTÃO DE PROJETOS

A gestão de projetos é a forma de realização de um projeto, em que a aplicação de técnicas de gestão durante o ciclo de vida do projeto ajuda a atingir objetivos específicos. Segundo a AFNOR, a gestão de projetos abrange todas as ferramentas, técnicas e métodos que permitem ao gestor liderar, coordenar e harmonizar as várias tarefas realizadas no âmbito do projeto de modo que atenda às necessidades explícitas e implícitas para as quais foi realizado.

Mas porque a gestão de projetos é importante? Porque nada é feito sem primeiro desenvolver um plano de projeto, e nenhum plano de projeto é executado sem o ambiente ou processos corretos. gerenciamento de projetos é, então, a ação que ajuda a criar e executar esse plano de projeto. Ele aplica competências gerenciais e de relacionamento interpessoal ao processo de execução de um projeto, da concepção à conclusão, de acordo com os requisitos estabelecidos.

O gerenciamento de projetos também tem a virtude de gerenciar o paradoxo entre o conhecimento do cliente e a capacidade de ação:

- No início do projeto, a equipe do projeto tem capacidade significativa de ação, mas pouco conhecimento.
- Então, e ao longo do projeto, o conhecimento aumenta e a capacidade de ação diminui.

Na verdade, o gerenciamento de projetos permite trazer o máximo de conhecimento no início do projeto quando você tem uma grande capacidade de ação. Tudo se resume a pensar com cuidado antes de começar.

3.2 GESTÃO DE PROJETOS EM CIÊNCIA DE DADOS

Para as empresas, a análise de dados se tornou um importante fator de crescimento. Mas desenvolver um projeto de dados é um processo complexo que requer o cumprimento de certos critérios para que se transforme em um sucesso real.

Muitos projetos de dados centram-se na implementação de soluções tecnológicas (Data Lake, instalação de clusters Hadoop, utilização de base de dados NoSQL ...) sem sequer se preocupar com a sua finalidade. Os investimentos são, de fato, focados em TI e não em linhas de negócios e geram pouco input de negócios.

Todos esses projetos não são necessariamente uma perda total, pois contribuem pelo menos para o aumento das habilidades tecnológicas das equipes de TI e do departamento de TI, mas geram pouco valor para a empresa.

Todos esses projetos não são necessariamente uma perda total, pois contribuem pelo menos para o aumento das habilidades tecnológicas das equipes de TI e do departamento de TI, mas geram pouco valor para a empresa.

4. MÉTODOS DE GESTÃO DE PROJETOS

Para ter sucesso com seu projeto de ciência de dados dentro do prazo e do orçamento, você e sua equipe precisam ser organizados e eficientes. O segredo? Siga um dos métodos de gerenciamento de projeto existentes para ajudá-lo a organizar seu projeto de forma simplificada e estruturada.

As metodologias de gerenciamento de projetos ajudam você a realizar todas as etapas de seu projeto, do planejamento à implementação, visando eficiência e lucratividade.

A escolha de uma metodologia para liderar um projeto permite que todas as partes interessadas trabalhem juntas de forma eficaz, seguindo regras claramente definidas. Neste artigo, iremos abordar três perspectivas de gestão de projetos: os métodos tradicionais, os métodos ágeis e os métodos emergentes.

4.1 MÉTODOS TRADICIONAIS

Os métodos clássicos são os métodos mais amplamente usados em gerenciamento de projetos. Esses métodos também são chamados de "cascata" porque cada etapa deve ser concluída para passar para a próxima.

Ao aplicar esta metodologia, a equipe do projeto segue as especificações de forma literal e trabalha em todo o projeto até a sua entrega. Não há interação com o cliente que receberá seu projeto assim que for concluído.

Tudo deve ser planejado. A equipe se compromete com um cronograma preciso e define todas as tarefas a serem realizadas. A principal desvantagem? Não há espaço para mudanças e o inesperado, então é melhor fazer tudo certo da primeira vez.

Para que esses métodos funcionem corretamente, uma boa comunicação entre os membros da equipe é essencial.

Um dos métodos mais usados é o **Waterfall**. O método é conhecido como um método tradicional de gerenciamento de projetos. Entre as várias metodologias existentes, este é o método mais utilizado, é um modelo de negócios linear que divide os processos de desenvolvimento em fases sucessivas do projeto.

Os requisitos de todas as partes interessadas são coletados no início e, em seguida, um cronograma do projeto é elaborado para atender a esses requisitos de maneira ordenada, sem possibilidade de retrocesso.

As principais etapas do método waterfall são:

- Requisitos: análise e expressão das necessidades do cliente;

- Análise: desenvolvimento de especificações e definição de especificações funcionais;
- Desenho do projeto: ou seja, seu planejamento;
- Implementação: produção do produto de acordo com as especificações;
- Validação (teste): o produto é testado pela equipe do projeto e verificado pelo diretor;
- Deploy: o produto é validado.

4.2 MÉTODOS AGEIS

Mais eficientes e menos rígidos do que os métodos tradicionais, os métodos ágeis colocam as necessidades do cliente no centro das prioridades do projeto. Eles oferecem maior flexibilidade e melhor visibilidade no gerenciamento de projetos, o que permite que a equipe seja mais sensível às expectativas do cliente.

A metodologia ágil não surgiu como um movimento espontâneo. Na verdade, sua criação foi pensada de maneira meticulosa, em uma resposta às necessidades dos desenvolvedores de software. Na década de 1990, quando aconteceu o primeiro boom da Internet, o desenvolvimento de softwares era baseado no modelo "waterfall" ou cascata.

No começo de 2001, um grupo de 17 desenvolvedores reconhecidos se juntou em Utah, nos EUA, para discutir maneiras de desenvolvimento mais leves com base em suas experiências. Eles assinaram um documento chamado "Manifesto para o Desenvolvimento Ágil de Software".

Este manifesto foi adaptado para o contexto de ciência de dados, no livro **Agile Data Science** de Jurney (2017). Os princípios de agilidade para ciência de dados são:

- **Iterar, iterar, iterar: tabelas, gráficos, relatórios, previsões**

Os insights demoram a surgir, não tem jeito. Não é a primeira consulta aos dados nem o primeiro gráfico formado que irá indicar a oportunidade de melhoria. A iterações fazem parte dos processos de Data Science, assim como na extração e na visualização de dados.

Portanto, a cada nova etapa de coleta de dados, seja em medições, reunião de acompanhamento de projeto ou atualização de cronogramas busque realizar novas análises, diferentes visualizações, segmentações de dados, filtros, busque correlações entre indicadores e avalie a necessidade de novos dados ou retirada de dados desnecessários.

- **Envie a produção intermediária. Mesmo experimentos fracassados têm resultados**

Se iterações são atividades essenciais, é comum nas metodologias ágeis os Sprints serem finalizados sem estarem completos, o "feito é melhor que perfeito".

Não tem como entregar uma etapa de um projeto, colher feedbacks e corrigir depois como um software ou um aplicativo, é claro. Porém, se o cronograma for organizado de forma a reduzir os lotes de entregas é possível obter informações acerca de produtividade, custos, materiais e restrições com maior frequência.

Entregamos um sprint, obtemos os feedbacks e temos a matéria prima para analisar e tomar decisões estratégias acerca da continuidade do projeto, os dados. Isto reflete em reuniões semanais mais ricas e eficazes.

- **Entregue etapas intermediarias**

A iteração é o ato essencial na construção de aplicativos analíticos, o que significa que muitas vezes nos encontramos no final de um sprint com coisas que não estão completas. Se não estivéssemos entregando resultados incompletos ou intermediários no final de um sprint, muitas vezes acabaríamos não enviando nada. E isso não é ágil. Podemos definir como um «ciclo da morte", onde tempo infinito pode ser desperdiçado aperfeiçoando coisas que ninguém deseja.

Bons sistemas são auto documentados e, na ciência de dados Ágil, documentamos e compartilhamos os recursos incompletos que criamos enquanto trabalhamos. Dedicamos todo o trabalho ao controle de origem. Compartilhamos esse trabalho com nossos colegas de equipe e, o mais rápido possível, com os usuários finais.

- **Conduza experimentos, não tarefas**

No campo da engenharia de software, um gerente de produto atribui a um desenvolvedor um gráfico a ser implementado durante um sprint. O desenvolvedor traduz este gráfico em uma consulta GROUP BY SQL e cria uma página da web para ele. Missão cumprida? Falso. Os gráficos especificados dessa maneira provavelmente não terão valor. A ciência de dados difere da engenharia de software por ser uma ciência e uma engenharia.

Para qualquer tarefa, temos que iterar para obter algo relevante, e essas iterações podem ser chamadas de experimentos, na melhor das hipóteses. Gerenciar uma equipe de ciência de dados significa supervisionar vários experimentos simultâneos, em vez de distribuir tarefas.

Os recursos certos (tabelas, gráficos, relatórios, previsões) aparecem como artefatos de análise exploratória de dados, portanto, precisamos pensar mais em termos de experiências do que de tarefas.

- **Escutar os dados**

O que é possível é tão importante quanto o que está planejado. O que é fácil e o que é difícil é tão importante saber quanto o que se deseja. No desenvolvimento de aplicativos de software, há três perspectivas a serem consideradas: perspectivas do cliente, do desenvolvedor e do negócio.

No desenvolvimento de aplicações analíticas, existe outra perspectiva: a dos dados. Sem entender o que os dados "têm a dizer" sobre uma característica, o product owner não pode fazer um bom trabalho. A opinião de dados deve sempre ser incluída nas discussões do produto, o que significa que deve ser baseada na visualização por meio de análise exploratória de dados como parte do aplicativo interno que se torna o foco.

- **Respeitar a pirâmide de valor dos dados**

A pirâmide de valor de dados (figura abaixo) é uma pirâmide de cinco níveis modelada na hierarquia de necessidades de Maslow. Expressa a quantidade crescente de valor criado pelo refinamento de dados brutos na forma de tabelas e gráficos, seguidos de relatórios e previsões, tudo com o objetivo de possibilitar novas ações ou melhorar as ações existentes:

- A primeira camada representa o aumento do valor agregado conforme os dados são refinados, desde os dados brutos coletados nos pipelines dos processos e representados por registros (CALDINI,2020)
- A segunda camada representa o processo de tratamento e refinamento dos dados, geralmente com gráficos e análises exploratórias dos dados
- A camada de relatórios permite uma exploração imersiva dos dados, onde podemos realmente discuti-los e conhecê-los.
- A camada de previsões é onde você cria mais valor, mas para criar boas previsões você tem que fazer engenharia de recursos, que os níveis mais baixos abrangem e facilitam. É nesta etapa que o Business Intelligence mostra o seu potencial, possibilitando as estruturas, links e interações entre os dados armazenados na organização (CADINI, 2020).
- A camada de previsões, é a etapa que se cria mais valor com a utilização de técnicas de Machine Learning, Data Mining.
- O último nível, o das ações, é onde a mania por inteligência artificial (IA) se manifesta. Se a sua intuição não permite realizar uma nova ação ou melhorar uma já existente, ela tem pouco valor.

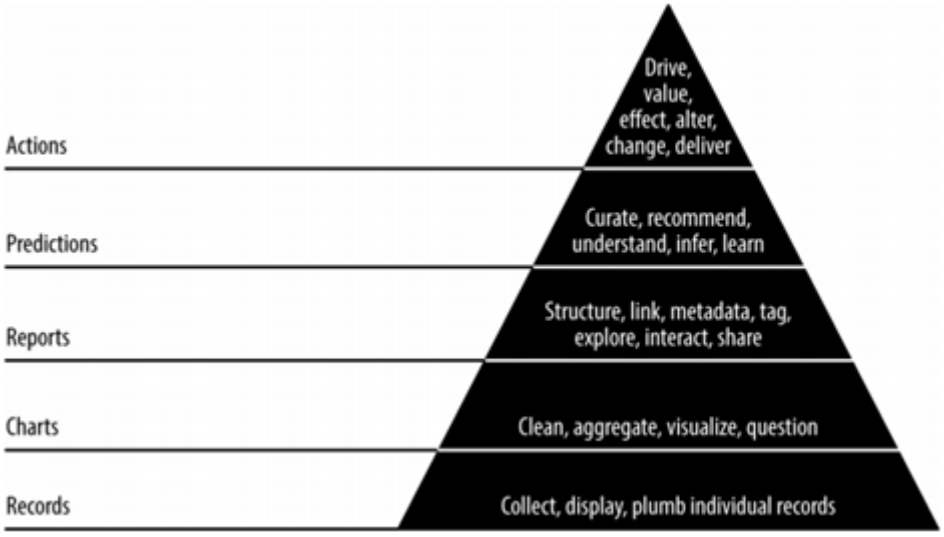


FIGURA 1: Pirâmide de Valor de Dados (Jurney, 2017)

A pirâmide de valor dos dados dá estrutura ao nosso trabalho. A pirâmide é algo para se manter em mente, não uma regra a seguir. Às vezes, pulamos etapas, às vezes trabalhamos ao contrário.

- **Obtenha meta. Descreva o processo, não apenas o estado final.**

Se não pudermos entregar bons produtos com facilidade em um cronograma comparável ao desenvolvimento de um aplicativo normal, o que faremos? Se não entregamos, não somos Ágil. Para resolver este problema, na ciência de dados Agil, nós "vamos para meta". A ênfase está em documentar o processo analítico em oposição ao estado final ou produto que estamos procurando.

Isso nos permite ser ágeis e fornecer conteúdo intermediário enquanto escalamos iterativamente a pirâmide de valor dos dados para encontrar o caminho crítico para um produto incrível. Então, de onde vem o produto? A partir da paleta, criamos e expandimos documentando nossa análise exploratória de dados.

Esses sete princípios trabalham juntos para alimentar o método de ciência de dados Ágil. Eles servem para estruturar e documentar o processo de análise exploratória de dados e transformá-lo em aplicações analíticas.

A metodologia Ágil Scrum é, sem dúvida, o método mais usado nos dias de hoje, principalmente porque pode ser integrado a outros métodos ágeis com facilidade. O Scrum é um framework dentro do qual pessoas podem tratar e resolver problemas complexos e adaptativos, enquanto produtiva e criativamente entregam produtos com o mais alto valor possível (Sutherland, Schwaber, 2020).

Scrum é baseado na divisão de um projeto em "caixas de tempo", chamadas de sprints ("picos de velocidade"). Os sprints podem durar de algumas horas a um mês (com um sprint médio de duas semanas). Cada sprint começa com uma estimativa seguida pelo planejamento operacional. O sprint termina com uma demonstração do que foi concluído. Antes de iniciar um novo sprint, a equipe realiza uma retrospectiva. Esta técnica analisa o andamento do sprint concluído, a fim de aprimorar suas práticas. O fluxo de trabalho da equipe de desenvolvimento é facilitado por sua auto-organização, portanto, não haverá gerente de projeto.

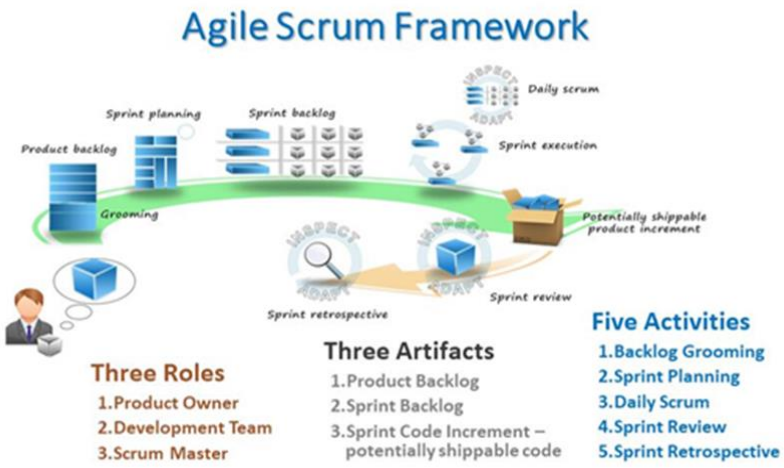


FIGURA 2: Etapas Scrum (Fonte: https://medium.com/the-liberators/refresh-the-purpose-of-the-scrum-framework-9e4bceb25499, 2019)

O Scrum é muito rígido para um projeto de ciência de dados com tarefas circulares, porque a natureza exploratória da análise de dados não é adequada para o planejamento restrito de entregas. **Kanban**, por outro lado, nos dá flexibilidade para mudar a próxima tarefa que vamos completar.

Kanban é uma ferramenta visual para tomar decisões rápidas. Esta ferramenta garante uma melhor colaboração e comunicação entre os colaboradores, mas também um bom fluxo de informação sobre as tarefas a realizar e quando as realizar. As tarefas são representadas por cartões, etiquetas ou post-its que são movidos de uma coluna para outra, dependendo do status da tarefa.

FIGURA 3: Kanban Board (Fonte: https://www.amalo-recrutement.fr/blog/kanban/, 2021)

O objetivo do Kanban é, portanto, se adaptar constantemente às necessidades do cliente. Objetivo: limitar o risco de superprodução e desperdício, mas também reduzir atrasos e custos. Qualificado como um método ágil da mesma forma que o Scrum, o Kanban preconiza a

visualização dos fluxos de trabalho através de um quadro (denominado quadro Kanban), possibilitando priorizar e monitorar o andamento das tarefas a serem realizadas.

O kanban board (Kanbanzine, 2021), é uma ferramenta para visualização e um dos componentes chave do método Kanban. Originalmente construído usando um quadro em branco, ele é dividido em colunas e raias. Cada coluna representa uma etapa do fluxo de trabalho e as raias separam diferentes tipos de atividades. Quando uma tarefa entra no seu fluxo de trabalho, ele é colocado em um cartão Kanban que passa por cada coluna do quadro. É por isso que você pode chamá-lo de quadro de tarefas Kanban. As principais colunas do kanban board são:

- **A fazer**: essas são as tarefas a serem realizadas.
- **Em andamento**: esta coluna agrupa as tarefas em andamento.
- **A testar**: aqui encontram-se as tarefas a testar, aquelas que aguardam validação do cliente.
- **Concluído**: Finalmente, na última coluna estão as tarefas concluídas.

O método kanban se baseia em 4 princípios:

- Comece fazendo o que você sabe fazer e com o que você já tem: o método Kanban usa os processos já implantados e incentiva a melhoria dos processos já implantados depois disso.
- Aceitar a aplicação de mudanças graduais: a equipe deve concordar em melhorar o sistema em vigor gradualmente, isso deve ser feito aos poucos
- Respeitar o atual processo, funções, responsabilidades e títulos: para evitar a pressa das equipes, será necessário respeitar os papéis, responsabilidades e títulos de cada um.
- Liderança em todos os níveis: sejam funcionários ou gerentes, todos os atores da cadeia produtiva que desejam implementar a melhoria contínua devem ser incentivados.

Kanban pode ser eficaz para ciência de dados. Seus processos fluidos e menos rigorosos fornecem aos cientistas de dados maior flexibilidade para executar seu trabalho sem ter que cumprir prazos constantes.

Como outras abordagens ágeis, o trabalho é dividido em pequenos incrementos, o que permite iterações rápidas e entrega contínua. Kanban fornece uma estrutura que é mais do que muitas equipes de ciência de dados têm atualmente (Saltz, Shamshurin & Crowston, 2017).

- Comece fazendo o que você sabe fazer e com o que você já tem: o método Kanban usa os processos já implantados e incentiva a melhoria dos processos já implantados depois disso.
- Aceitar a aplicação de mudanças graduais: a equipe deve concordar em melhorar o sistema em vigor gradualmente, isso deve ser feito aos poucos
- Respeitar o atual processo, funções, responsabilidades e títulos: para evitar a pressa das equipes, será necessário respeitar os papéis, responsabilidades e títulos de cada um.
- Liderança em todos os níveis: sejam funcionários ou gerentes, todos os atores da cadeia produtiva que desejam implementar a melhoria contínua devem ser incentivados.

Kanban pode ser eficaz para ciência de dados. Seus processos fluidos e menos rigorosos fornecem aos cientistas de dados maior flexibilidade para executar seu trabalho sem ter que cumprir prazos constantes. Como outras abordagens ágeis, o trabalho é dividido em pequenos incrementos, o que permite iterações rápidas e entrega contínua. Kanban fornece uma estrutura que é mais do que muitas equipes de ciência de dados têm atualmente (Saltz, Shamshurin & Crowston, 2017).

4.3 MÉTODOS EMERGENTES

Nesta etapa, iremos conhecer dois métodos: crisp -dm desenvolvido pela IBM e Team Data Science Process desenvolvido pela Microsoft.

Cross Industry Standard Process for Data Mining (CRISP-DM) é um modelo de processo de mineração de dados que descreve uma abordagem comumente usada para resolver problemas no campo [pouco claro] de análise, mineração e ciência de dados (Chapman, 2000).

A metodologia foi criada há pouco mais de 20 anos, pela necessidade dos profissionais de Data Mining. Apesar de existir uma série de ferramentas capazes de nortear esses profissionais, quando o assunto é Big Data e o seu grande volume de dados, elas deixam a desejar. O CRISP DM surgiu justamente para atender aos projetos que estão diretamente envolvidos com o processamento e a análise de um grande volume de dados.

Com base na experiência prática e real de como as pessoas conduzem projetos de mineração de dados, o método CRISP-DM é um processo que compreende as seguintes etapas (figura abaixo).

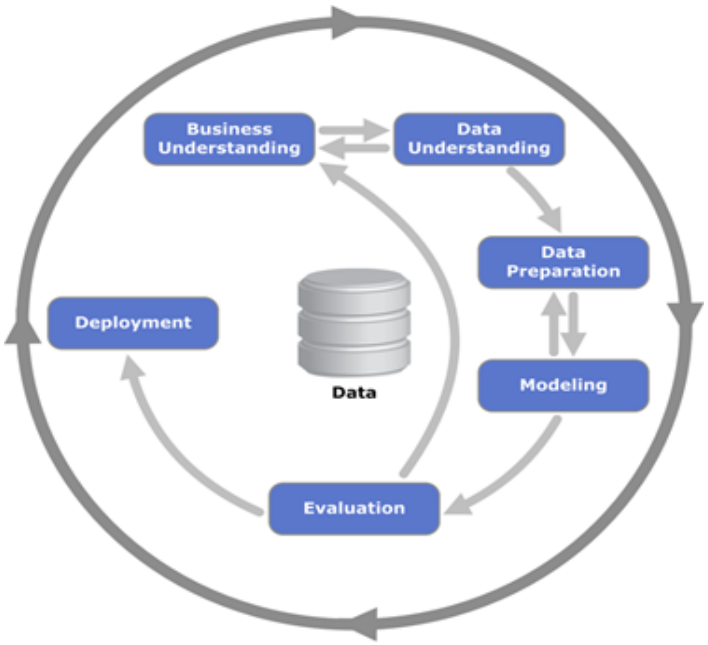


FIGURA 4 : Etapas Crisp-dm (IBM,2021)

As seis etapas representadas aqui são:

- **Business Understanding:** concentra-se em compreender os objetivos e requisitos do projeto de uma perspectiva de negócios e, em seguida, traduzir essas informações em uma definição de problema de ciência de dados.
- **Data Understanding**: concentra-se em coletar e familiarizar-se com os dados; isso é relevante para identificar problemas de qualidade de dados, descobrir os primeiros insights sobre os dados e formular hipóteses.
- **Data Preparation**: visa transformar os dados brutos em um conjunto de dados final que pode ser usado como entrada para técnicas de modelagem (por exemplo, algoritmos de aprendizado de máquina).
- **Modeling**: envolve a aplicação de diferentes técnicas de modelagem ao conjunto de dados para gerar um conjunto de modelos candidatos.
- **Evaluation**: uma vez que os modelos foram construídos, eles precisam ser testados para garantir que generalizem contra dados invisíveis e que todos os principais objetivos de negócios tenham sido considerados (por exemplo, o modelo final precisa ser justo, interpretável por humanos e atingir uma precisão X% maior do que a solução atual do cliente). O resultado dessa etapa é o modelo campeão.
- **Deployment:** o modelo campeão é implantado na produção para que possa ser usado para fazer previsões sobre dados invisíveis. Todas as etapas de preparação de dados são incluídas para que o modelo trate os novos dados brutos da mesma maneira que durante o desenvolvimento do modelo.

Após implantado o ciclo do CRISP DM, é possível ter análises em tempo real conforme a situação e o cenário vão mudando, possibilitando mudanças imediatas e personalizadas para cada momento. A agilidade na tomada de decisões e a resolução de problemas, com certeza, são vantagens competitivas importantes.

Outro método bastante utilizado é o Microsoft Team Data Science Process (TDSP). O TDSP é uma metodologia abrangente que define funções, processos e modelos que são amplamente inspirados no CRISP-DM e Scrum. Lançado em 2016, o TDSP foca nos aspectos do ágil focado em entregar soluções de ciência de dados de forma eficiente. O TDSP ajuda a aprimorar a colaboração e o aprendizado da equipe sugerindo como as funções de equipe funcionam melhor em conjunto (Microsoft, 2021).

Como cada projeto e equipe de ciência de dados são diferentes, cada ciclo de vida específico de ciência de dados é diferente. No entanto, a maioria dos projetos de ciência de dados tende a seguir o mesmo ciclo de vida geral.

O TDSP tem os seguintes componentes principais:

- Uma definição de ciclo de vida da ciência de dados
- Uma estrutura de projeto padronizada
- Infraestrutura e recursos recomendados para projetos de ciência de dados

- Ferramentas e utilitários recomendados para execução do projeto
- Aqui está uma representação visual do ciclo de vida do Team Data Science Process.

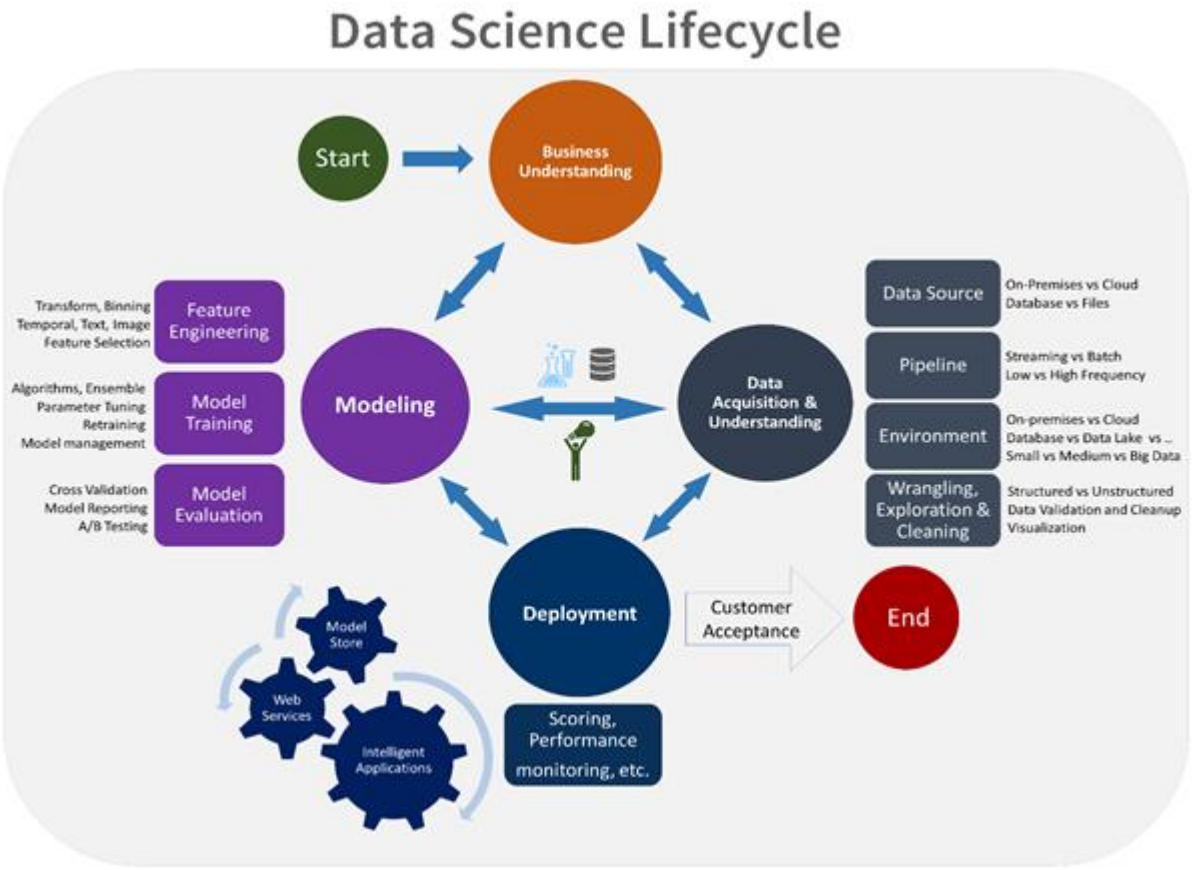


FIGURA 5: ciclo de vida do TDSP (Microsoft, 2021)

O ciclo de vida do TDSP é composto por cinco estágios principais executados de forma iterativa. Esses estágios incluem:

● **Business understanding**: Para construir um modelo de negócios de sucesso, é muito importante primeiro entender o problema de negócios que o cliente está enfrentando. Para te auxiliar nesta etapa, você pode usar as técnicas: 5-Why, Analise Swot, Gemba Walk, diagrama de Ishikawa, etc.

● **Data acquisition and understanding**: Depois de obter clareza sobre a definição do problema, precisamos coletar dados relevantes para dividir o problema em pequenos componentes. O objetivo desta etapa é produzir um conjunto de dados limpo e de alta qualidade cuja relação com as variáveis de destino seja compreendida.

● **Modeling**: Estabeleça um relacionamento entre entradas e saídas, iterando os dados e o algoritmo para atingir o valor comercial.

● **Deployment:** Todos os projetos de ciência de dados devem ser implantados no mundo real. A implantação pode ser por meio de um aplicativo Android ou iOS. Os modelos de aprendizado de máquina podem ter que ser registrados antes da implantação, porque os cientistas de dados podem favorecer a linguagem de programação Python, mas o ambiente de produção oferece

suporte a Java. Depois disso, os modelos de aprendizado de máquina são implantados primeiro em um ambiente de pré-produção ou teste antes de realmente implantá-los na produção.

- **Customer acceptance**: Finalize as entregas do projeto: confirme se o pipeline, o modelo e sua implantação em um ambiente de produção atendem aos objetivos do cliente.

O diagrama a seguir fornece uma visão de grade das tarefas (em azul) e artefatos (em verde) associados a cada estágio do ciclo de vida (no eixo horizontal) para essas funções (no eixo vertical).

O ciclo de vida do TDSP é modelado como uma sequência de etapas iteradas que fornecem orientação sobre as tarefas necessárias para usar modelos preditivos. Você implementa os modelos preditivos no ambiente de produção que planeja usar para construir os aplicativos inteligentes. O objetivo deste método é continuar a mover um projeto de ciência de dados em direção a um ponto final de engajamento claro. A ciência de dados é um exercício de pesquisa e descoberta.

A capacidade de comunicar tarefas para sua equipe e seus clientes usando um conjunto bem definido de artefatos que empregam modelos padronizados ajuda a evitar mal-entendidos. O uso desses modelos também aumenta a chance de conclusão bem-sucedida de um projeto complexo de ciência de dados.

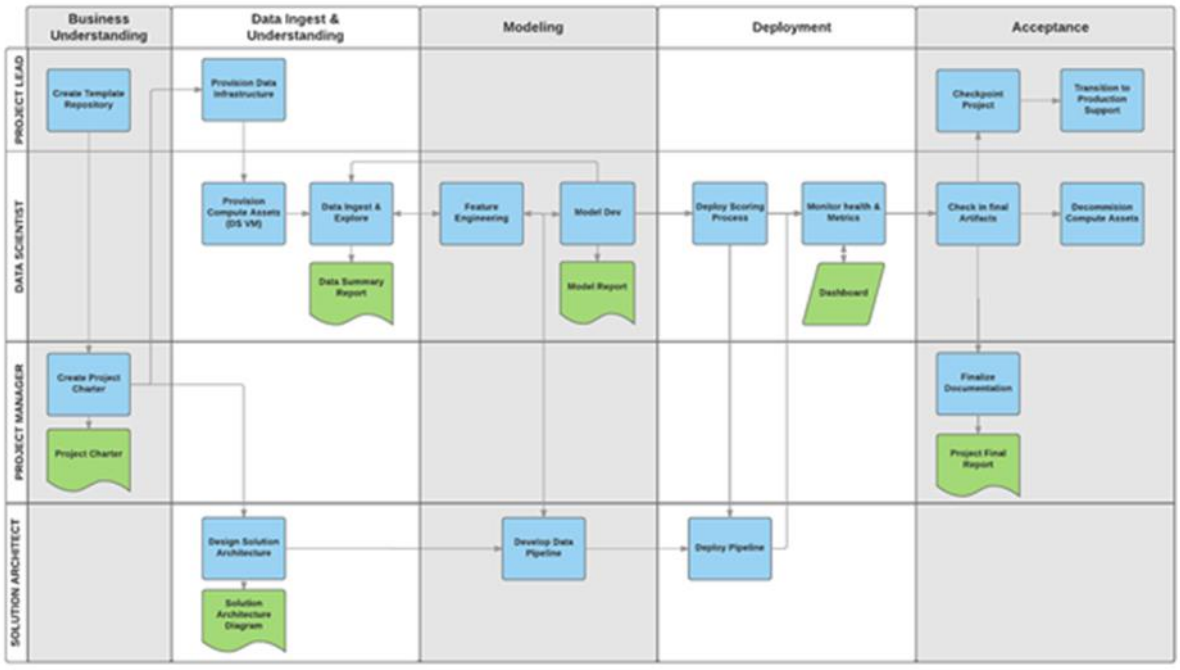


FIGURA 6: Tasks (Microsoft, 2021)

O ciclo de vida do TDSP é modelado como uma sequência de etapas iteradas que fornecem orientação sobre as tarefas necessárias para usar modelos preditivos. Você implementa os

modelos preditivos no ambiente de produção que planeja usar para construir os aplicativos inteligentes. O objetivo deste método é continuar a mover um projeto de ciência de dados em direção a um ponto final de engajamento claro. A ciência de dados é um exercício de pesquisa e descoberta.

A capacidade de comunicar tarefas para sua equipe e seus clientes usando um conjunto bem definido de artefatos que empregam modelos padronizados ajuda a evitar mal-entendidos. O uso desses modelos também aumenta a chance de conclusão bem-sucedida de um projeto complexo de ciência de dados.

Conforme vimos anteriormente, existem diferentes abordagens e metodologias que podem ser aplicadas para gerenciar projetos de Data Science. De fato, é necessário compreender suas especificidades, de tal modo que se possa avaliar qual abordagem é mais eficiente de acordo com o cenário em que se pretende aplicar.

- Bibliografia

AFNOR. **Gestion de Projets**. https://competences.afnor.org/gammes/gestion-de-projets 2021

Chapman, P. **CRISP-DM 1.0: Step-By-Step Data Mining Guide**, 2000.
IBM. Crisp-DM. https://www.ibm.com/docs/fr/spss-modeler/SaaS?topic=dm-crisp-help-overview 2021

ISO 10006:2017. **Quality management — Guidelines for quality management in projects**
JURNEY, Russel. Agile Data Science. 2017

Kanban Board. https://www.amalo-recrutement.fr/blog/kanban/ 2021

SALTZ, Shamshurin, Crowston. **Comparing Data Science Project Management Methodologies via a Controlled Experiment**. 2017

Scrum Framework. https://medium.com/the-liberators/refresh-the-purpose-of-the-scrum-framework-9e4bceb25499 2019

O que é Kanban? https://kanbanize.com/pt/recursos-kanban/primeiros-passos/o-que-e-quadro-kanban. 2021

What is the Team Data Science Process? https://docs.microsoft.com/en-us/azure/architecture/data-science-process/overview. 2021

5. O QUE É APRENDIZAGEM?

Aprendizagem é um processo de mudança de comportamento obtido através da experiência construída por fatores emocionais, neurológicos, relacionais e ambientais. Portanto podemos dizer que “aprender” é o resultado da interação entre estruturas mentais e o meio ambiente.

Esse processo - aprendizagem - pode ser medido através das curvas de aprendizagem, que mostram a importância da repetição de certas predisposições fisiológicas, de “tentativa e erro” e de períodos de descanso.

Há um número grande de formas de aprendizagem, dentre as quais podemos citar a adaptação, a correção, a otimização, a representação e a interação. A aprendizagem humana está ligada à inferência de regras por trás dos padrões e segundo as teorias conexionistas. Está relacionada ao número de sinapses feitas pelos neurônios, isso quer dizer que “aprender significa fazer novas conexões”.

FIGURA 7: Nuvem de Palavras associadas a “Aprendizagem”

5.1 O QUE É O APRENDIZADO DE MÁQUINA

O aprendizado de máquina, em inglês machine learning (M.L.), se trata de uma tecnologia que utiliza Inteligência Artificial como base e algoritmos que aprendem interativamente por meio de dados coletados em suas interações. Com o uso do M.L., nós ensinamos a máquina a aprender as coisas e, assim, realizar determinadas tarefas sozinha.

Imaginemos, por exemplo, que um aplicativo de M.L. tem o objetivo de identificar fotos de frutas em um banco de dados. Neste exemplo, o aplicativo de M.L. tem o objetivo de identificar e agrupar as frutas da mesma variedade.

O que aplicativo faz:

- Com base nos dados introduzidos, a aplicação de ML analisa os dados.
- Em seguida, tenta identificar padrões, como por exemplo tamanhos, formatos e cores.
- Baseado nos padrões identificados irá segregar os diferentes tipos ou variedades de frutas, criando grupos específicos destas variedades.
- Por último, ele mantém o controle de todas as decisões tomadas no processo para garantir que está aprendendo

5.2 ABORDAGENS DE APRENDIZADO DE MÁQUINA

Existem diferentes tipos de abordagens, ou tipos de aprendizado de máquina, e que são tradicionalmente divididas em três categorias amplas (figura abaixo).

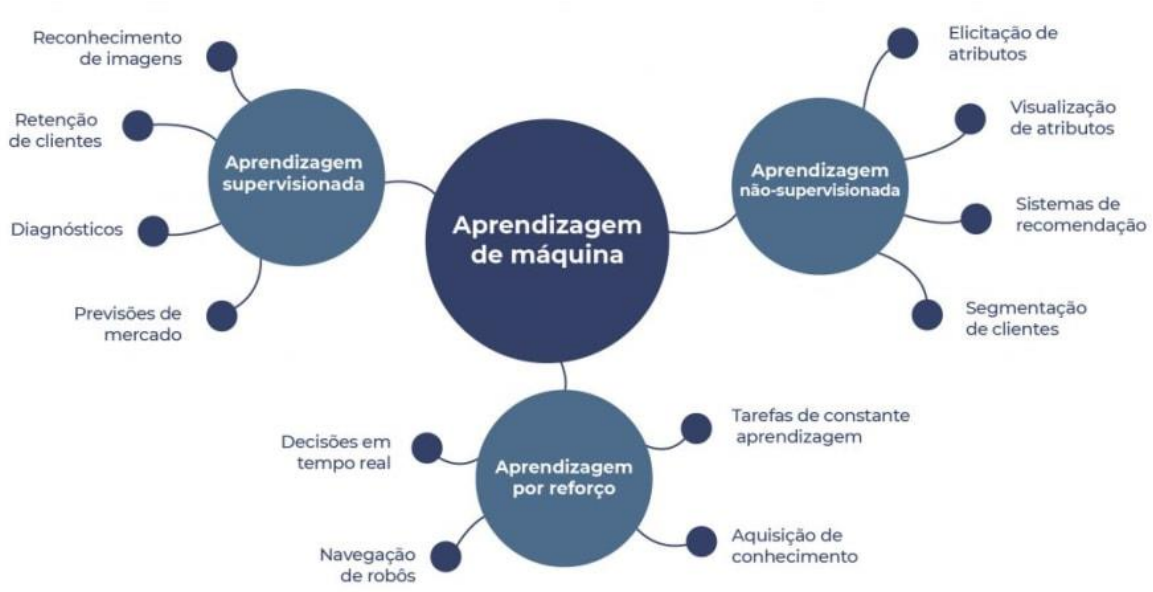


FIGURA 8: Tipos de aprendizagem de máquina
(Fonte: https://dev.to/beatrizmaiads/tipos-de-aprendizado-de-maquina-3-5d66)

5.2.1 Aprendizagem supervisionada

Nesta abordagem dizemos ao computador o que é cada entrada (qual o rótulo) e ele irá aprender quais características daquelas entradas fazem elas serem o que são.

Ocorre quando o modelo aprende a partir de resultados pré-definidos. Ao computador são apresentados exemplos de entradas e suas saídas desejadas, e o objetivo é que ele aprenda uma regra geral que mapeia entradas em saídas.

O algoritmo de aprendizagem supervisionada é mais comumente utilizado, com base de dados previamente rotulados, sendo possível fazer previsões futuras.

Neste tipo de aprendizagem, trabalhamos com dois tipos de problemas: classificação e regressão.

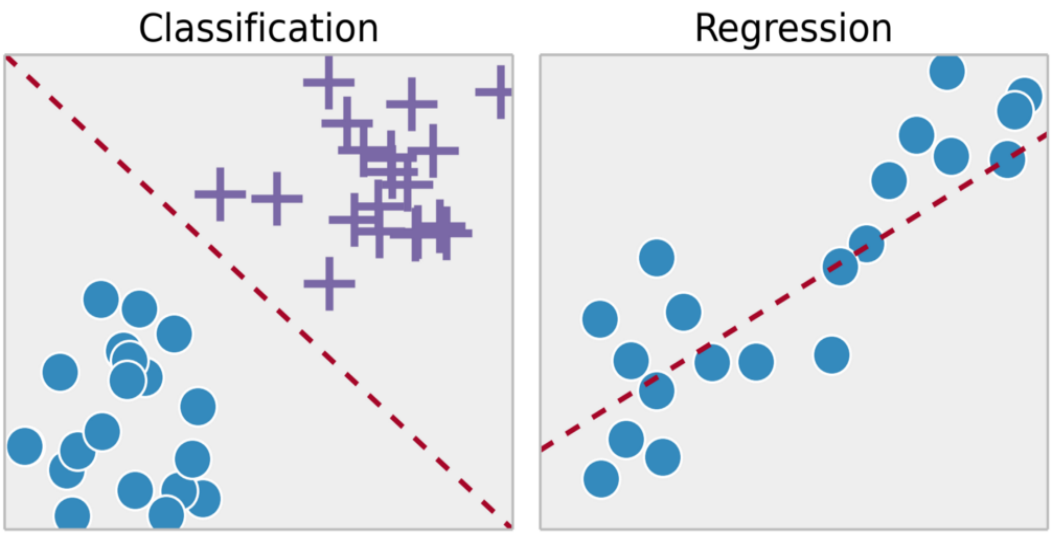


FIGURA 9: Tipos de aprendizagem supervisionada
(Fonte: https://medium.com/opensanca/aprendizagem-de-maquina-supervisionada-ou-n%C3%A3o-supervisionada-7d01f78cd80a)

Imagine um programa que usa Machine Learning para identificar fotos de cachorros. Vou usar dois dados de entrada para treiná-lo:

- Imagem: fotos variadas, algumas de cachorros, e outras não.
- Booleano: um booleano que indica se a foto é ou não de um cachorro.

O treinamento ocorre quando o programa "vê" uma imagem junto com sua resposta, ou seja, esta imagem é de um cachorro? Imagine isto se repetindo para milhares de imagens diferentes. Chega um momento em que o programa identifica as características que fazem uma imagem ser a imagem de um cachorro.

O exemplo acima é de classificação. Estamos classificando uma entrada entre dois tipos possíveis: cachorro ou não-cachorro.

Já a regressão ocorre quando o resultado é numérico. Por exemplo, um programa que calcula o valor que uma casa deveria ter com base em características como número de quartos, localização e ano de construção. Com base em exemplos de casas similares, o computador aprende a precificar novas casas.

5.2.2 Aprendizagem não supervisionada

A diferença da aprendizagem não supervisionada para a supervisionada é que aqui a aprendizagem ocorre com dados não rotulados, ou seja, não dizemos ao computador o que é aquela entrada.

Desta forma, na abordagem não supervisionada, nenhum rótulo é dado ao algoritmo de aprendizagem, deixando-o sozinho encontrar a estrutura em sua entrada. Neste caso, o algoritmo identifica os padrões e os organiza.

Diferentemente do supervisionado, o aprendizado não supervisionado não se baseia em dados rotulados e pode ter um objetivo em si mesmo, ou seja, o de descobrir padrões ocultos nos dados, ou ainda, um meio para um fim, o de descobrir a estrutura dos dados.

Existem muitos tipos de aprendizado não supervisionado, embora existam dois problemas principais que são frequentemente encontrados por um profissional: eles são agrupamentos que envolvem a localização de grupos na estimativa de dados e a densidade que envolve o resumo da distribuição de dados.

- **Clustering**: problema de aprendizado não supervisionado que envolve a localização de grupos de dados.
- **Estimativa de densidade**: problema de aprendizado não supervisionado que envolve resumir a distribuição de dados.

Um exemplo de algoritmo de clustering é k-Means, em que k se refere ao número de clusters a serem descobertos nos dados. Um exemplo de algoritmo de estimativa de densidade é o Kernel Density Estimation, que envolve o uso de pequenos grupos de amostras de dados intimamente relacionadas para estimar a distribuição de novos pontos no espaço do problema[1].

[1] RUSSEL, Stuart. **Artificial Intelligence : A Modern Approach**, 3ª edição, 2015

5.2.3 Aprendizagem por reforço

A aprendizagem por reforço é uma forma de ensinar ao computador qual ação priorizar dada uma determinada situação.

Segundo Sutton (2018) "aprender por reforço é aprender o que fazer – como mapear situações para ações – para maximizar um sinal de recompensa numérico. O aluno não é informado sobre quais ações executar, mas, em vez disso, deve descobrir quais ações geram mais recompensa, tentando-as."[2]

Podemos imaginar um programa responsável por dirigir um veículo autônomo.

Ele deve aprender a dirigir pelas ruas e transportar seus passageiros. Existem diversas formas de otimizar esta tarefa. Por exemplo, chegar ao destino no menor tempo possível e não causar nenhum acidente. Queremos que ele saiba o que fazer conforme o que ocorre à sua volta, e preferimos que ele demore um pouco mais do que causar um acidente, por exemplo. À medida que navega no espaço do problema, o programa recebe um feedback análogo às recompensas, que ele tenta maximizar.

Uma das questões relevantes é a definir a correta solução de ML para um dado problema, especificamente qual algoritmo deve ser utilizado. Esta definição tem algumas dependências importantes:

- Declaração do problema e sua natureza. É importante antes de tudo entender qual é o problema;
- Entender os dados quanto ao tamanho, qualidade e a natureza dos dados que serão utilizados;
- Complexidade do algoritmo.

Um outro exemplo de um problema de reforço é jogar um jogo em que o agente tem o objetivo de obter uma pontuação alta e pode fazer jogadas no jogo e receber feedback em termos de punições ou recompensas. Em muitos domínios complexos, o aprendizado por reforço é a única maneira viável de treinar um programa para executar em níveis altos.

Por exemplo, no jogo, é muito difícil para um ser humano fornecer avaliações precisas e consistentes de muitas posições, o que seria necessário para treinar uma função de avaliação diretamente a partir de exemplos. Em vez disso, o programa pode ser informado quando ganhou ou perdeu e pode usar essas informações para aprender uma função de avaliação que fornece estimativas razoavelmente precisas da probabilidade de vitória de qualquer posição.

[2] SUTTON, Richard. **Reinforcement Learning, second edition: An Introduction (Adaptive Computation and Machine Learning series)**, 2ª edição, 2018.

6. PROJETOS DESENVOLVIDOS DURANTE A MENTORIA

Durante estes 4 meses de mentoria (junho a setembro), foram desenvolvidos dois projetos. O primeiro consiste num projeto de análise de Churn e o segundo consiste na análise de dados de criptomoedas para predição do seu valor no mercado.

Antes de apresentar o código, irei fazer uma breve contextualização sobre estes dois temas.

6.1 Análise de Churn

Churn, numa definição mais generalista, é uma métrica que indica o número de clientes que cancelam em determinado período. Para calcular o churn, o que você precisa fazer é somar o número de clientes que cancelou seu produto/serviço no período analisado[3].

Assim, se a empresa inicia o ano com 90 clientes e, transcorridos 12 meses, termina com 79, significa que a sua taxa de cancelamento anual foi de 12,22% ou 1,02% mensal.

Esse acompanhamento é fundamental, pois permite entender o motivo das desistências e criar estratégias personalizadas de retenção para cada uma delas.

O Churn pode aumentar por diversos motivos, entre eles:

- Ausência de suporte ao cliente ou serviço deficitário;
- Falta de personalização do atendimento;
- Insatisfação com o produto ou serviço adquirido;
- Não oferecer benefícios na fidelização;
- Fatores externos imprevisíveis;
- Erro de segmentação.

Para analisar o Churn, podemos utilizar algumas métricas, tais como:

- Churn Rate: total de clientes cancelados / número total de clientes ativos do último mês
- Churn de Receita ou MRR Churn** MRR significa Monthly Recurring Revenue (receita recorrente mensal)
- MRR CHURN = SUM (MRR dos clientes cancelados). O MRR Churn também pode ser calculado em percentuais, representando o quanto isso equivale quando olharmos para o total de MRR do mês.
- MRR CHURN % = MRR Churn / MRR último mês. Algo interessante sobre a análise do MRR churn é o insight que ele possibilita sobre downgrades e upgrades, duas métricas muito importantes que não são medidas pelo churn rate, por exemplo.

[3] https://resultadosdigitais.com.br/blog/o-que-e-churn/ (2020)

Em muitos casos o MRR Churn acaba sendo mais importante que o churn rate, isso somente se as contas que estão saindo são contas de baixa receita. Porém, existe um contrapeso para compensar esse furo. É importante que seus maiores e mais importantes clientes continuem ativos e crescendo, o que significa trazer mais receita

A análise do Churn rate traz uma série de insights valiosos para a gestão, desde os seus impactos nas finanças da empresa até a qualidade dos serviços oferecidos.

O aumento da taxa de cancelamento indica a necessidade de mais investimentos para a captação de novos clientes, por exemplo, porque, para haver equilíbrio financeiro e projetar melhores resultados, é preciso ter, no mínimo, uma receita compatível com os custos operacionais.

É preciso ressaltar, também, que os custos para a captação de novos clientes são maiores que o valor investido para manter uma base fidelizada e, obviamente, para a gestão é mais vantajoso conquistar a sua receita por meios mais econômicos.

O Churn também pode apontar que as estratégias comerciais, o funil de vendas e as demais ferramentas de gestão de clientes não estão devidamente calibrados com o perfil do público-alvo. Como consequência, a falta de compatibilidade gerará cancelamentos.

O Churn se mostra fundamental para a empresa, pois analisado em conjunto com as demais métricas e indicadores do negócio pode oferecer diretrizes de gestão capazes de aumentar a eficiência comercial e, por consequência, os seus resultados.

- **Problema a ser analisado: Previsão e prevenção da rotatividade de clientes**

Prever e prevenir a rotatividade de clientes representa uma enorme fonte de receita potencial adicional para todos os negócios.

A rotatividade de clientes (também conhecido como atrito de clientes) se refere à quando um cliente (jogador, assinante, usuário etc.) cessa seu relacionamento com uma empresa. Os negócios online geralmente tratam um cliente como cancelado uma vez que um determinado período decorreu desde a última interação do cliente com o site ou serviço. O custo total da rotatividade inclui a receita perdida e os custos de marketing envolvidos na substituição desses clientes por novos. Reduzir a rotatividade é uma meta comercial importante de todos os negócios online.

A capacidade de prever que um determinado cliente corre um alto risco de abandono, embora ainda haja tempo para fazer algo a respeito, representa uma enorme fonte adicional de receita potencial para todos os negócios online. Além da perda direta de receita resultante do abandono do negócio por um cliente, os custos de aquisição inicial desse cliente podem ainda não ter sido cobertos pelos gastos do cliente até o momento. (Em outras palavras, adquirir aquele cliente pode ter sido na verdade um investimento perdedor.) Além disso, é sempre mais difícil e caro adquirir um novo cliente do que reter um cliente pagante atual.

Para ter sucesso na retenção de clientes que, de outra forma, abandonariam o negócio, os profissionais de marketing e especialistas em retenção devem ser capazes de (a) prever com antecedência quais clientes vão se agitar por meio da análise de churn e (b) saber quais ações de marketing terão maior retenção impacto em cada cliente em particular. Munido desse conhecimento, uma grande proporção da rotatividade de clientes pode ser eliminada.
Embora simples em teoria, as realidades envolvidas em atingir essa meta de "retenção proativa" são extremamente desafiadoras.

As técnicas de modelagem de previsão de rotatividade tentam entender os comportamentos e atributos precisos do cliente que sinalizam o risco e o momento da rotatividade do cliente. A precisão da técnica usada é obviamente crítica para o sucesso de qualquer esforço de retenção proativo. Afinal, se o profissional de marketing não souber que um cliente está prestes a se desligar, nenhuma ação será tomada em relação a esse cliente. Além disso, ofertas ou incentivos especiais com foco na retenção podem ser fornecidos inadvertidamente a clientes felizes e ativos, resultando em receitas reduzidas sem um bom motivo.

Infelizmente, a maioria dos métodos de modelagem de previsão de churn depende da quantificação do risco com base em dados e métricas estáticos, ou seja, informações sobre o cliente como ele existe agora. Os modelos de previsão de churn mais comuns são baseados em métodos estatísticos e de mineração de dados mais antigos, como regressão logística e outras técnicas de modelagem binária. Essas abordagens oferecem algum valor e podem identificar uma certa porcentagem de clientes em risco, mas são relativamente imprecisas e acabam deixando dinheiro na mesa.

- **Dicionário de dados: Conjuntos de dados de rotatividade de telecomunicações**

Cada linha representa um cliente; cada coluna contém os atributos do cliente. Os conjuntos de dados têm os seguintes atributos ou recursos:

• State: Estado
• Account length: Comprimento da conta
• Area code: Código de área
• International plan: Plano Internacional
• Voice mail plan: Plano de correio de voz
• Number vmail messages: Número de mensagens
• Total day minutes: Total de minutos por dia
• Total day calls: Total de chamadas diárias
• Total day charge: Cobrança diária total
• Total eve minutes: Total de minutos véspera
• Total eve calls: Total de chamadas á véspera
• Total eve charge: Carga total véspera
• Total night minutes: Total de minutos noturnos
• Total night calls: Total de chamadas noturnas
• Total night charge: Cobrança total da noite

• Total intl minutes: Total de minutos
• Total intl calls: Total de chamadas Internacionais
• Total intl charge: Cobrança Internacional total
• Customer service calls: Chamadas de atendimento ao cliente
• Churn: Cliente desistiu do serviço

6.2 ANÁLISE DE DADOS DE CRIPTOMOEDAS PARA PREDIÇÃO DO SEU VALOR NO MERCADO

Criptomoedas são moedas digitais que não existem fisicamente e não são emitidas por nenhum governo. Elas têm três principais funções:

- servir como meio de troca, facilitando as transações comerciais;
- reserva de valor, para a preservação do poder de compra no futuro;
- unidade de conta, quando os produtos são precificados e o cálculo econômico é realizado em função dela.

Uma desvantagem importante é que o preço das criptomoedas apresenta grande volatilidade (Infomoney, 2021)[4]. Seu preço varia conforme oferta e demanda. Além disso, elas ainda possuem baixo grau de aceitação nos estabelecimentos, por enquanto. Existe um risco de que os usuários apaguem ou percam suas moedas além de ser necessário se proteger da ação de hackers (Infomoney, 2021)[5].

Algoritmos de Machine Learning têm sido amplamente utilizados no mercado financeiro com o intuito de prever preço de ações (listei referências para ler depois no tópico). Os modelos mais utilizados para essa tarefa são métodos ensemble LSTMs e redes neurais recorrentes.

- **Problema a ser analisado:**

O problema a ser analisado aqui será prever o preço de criptomoedas baseado em dados disponíveis de mercado. Além disso, notícias e interações nas redes sociais têm impacto no preço e, portanto, serão adicionadas à análise.

Inicialmente o modelo será treinado com dados de valores de Bitcoin (BTC) por ser a mais famosa e depois de ter um modelo e estratégia clara funcionando, testar com outras criptomoedas. Outras criptomoedas: Bitcoin Cash (nova versão do Bitcoin), Ethereum (ETH), Tether (USDT), Ripple (XRP), Litecoin (LTC) etc.

[4] https://www.infomoney.com.br/ferramentas/criptomoedas/
[5] https://www.infomoney.com.br/guias/criptomoedas/

- **Dicionário de dados:**

• Os dados utilizados serão preços das ações num intervalo de tempo (à partir de 2015) da api yahoo finance

• Tweets de páginas e assuntos que possam ter impacto no modelo.

• Podem ser incorporadas notícias e outras fontes futuramente.

7. CÓDIGOS DOS PROJETOS : https://github.com/dsamentoria

- Análise de Churn

https://htmlpreview.github.io/?https://github.com/dsamentoria/analise_churn/blob/main/TelecomChurnNataliaMurad.html

https://github.com/dsamentoria/analise_churn/tree/main/Rafael%20Gallo%20-%20Modelo%20machine%20learning

- Criptomoedas

https://github.com/dsamentoria/bitcoin/blob/main/BitcoinV2NataliaMurad.ipynb
https://github.com/dsamentoria/bitcoin/tree/main/Gallo%20Raf

8. MENSAGEM DOS MENTORES

- **Anne Francine Martins**

Atuar como mentora foi um grande desafio e aprendizado. Apesar da diferença do fuso horario, das dificuldades encontradas no início da mentoria, conheci pessoas maravilhosas . Em julho, três pessoas do grupo desistiram, o que gerou um certo desânimo. Natalia foi a única pessoa do grupo inicial que continuou até o fim e quero parabeniza-la pela sua resiliência e entusiamo. Natalia é uma profissional incrível, com grande conhecimento em bioinformatica e partilhamos nosso amor por R.

Em agosto, Luis ingressou ao nosso grupo me auxiliando na mentoria após a dissolução do seu grupo. Neste mesmo período, tivemos a entrada do Rafael no nosso grupo como mentorado, dando o gás necessário para continuar o projeto.

Obrigada Natalia, Luis e Rafael por ser este grupo tão maravilhoso... A mentoria acabou, mas a amizade continua e que possamos fazer mais projetos juntos ....

- **Luiz Henrique de Oliveira Bueno**

No inicio enfrentei meu primeiro desafio , a saída dos quatro mentorados , havia resolvido continuar sozinho, mais não cumpriria o propósito do desafio. Foi então que recebi o convite da amiga e também mentora Anne, para somar forças e formar outro grupo. Então tive o privilégio de conhecer a Natália e Rafael como mentorados, e formamos uma excelente equipe.

Através desse desafio proposto pela DSA , tive a oportunidade de agregar conhecimento, e por em prática algumas experiências que podem ser perfeitamente trazidas a realidade em projetos de Data Science. A peça fundamental foi a equipe unida, um grande time que tive a honra de participar como mentor, e cumprimos nossa meta. O projeto acabou mais nossa amizade permanecera..... Obrigado Anne, Nátalia , Rafael