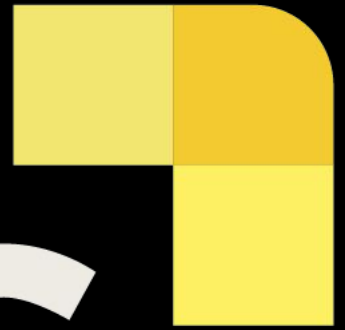


formação em dados



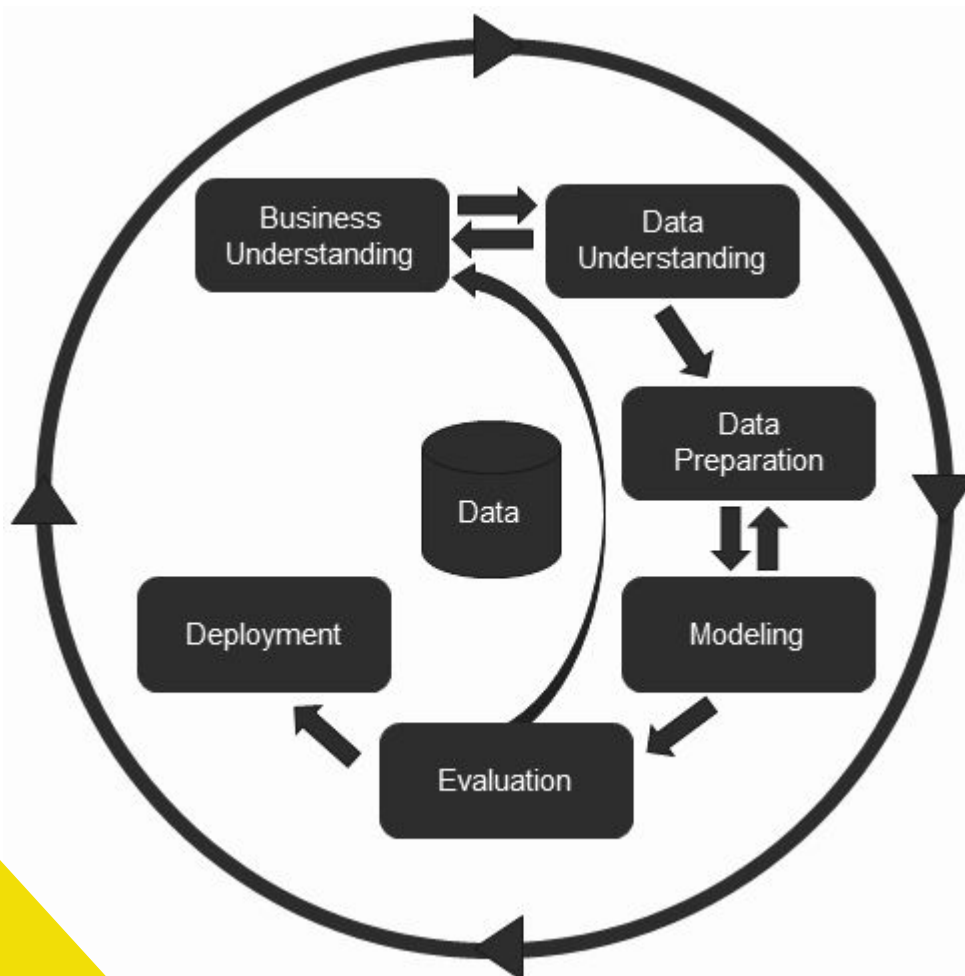
ETAPAS CRISP-DM

MATERIAL COMPLEMENTAR

CRISP-DM

Cross Industry Standard Process for Data Mining

Processo criado em meados de 1996, o CRISP DM é hoje a metodologia mais utilizada nas empresas para guiar projetos de data science. Tem como objetivo auxiliar os profissionais de mineração de dados principalmente, mas também trabalhando desde sua etapa inicial com o entendimento do problema de negócio.





Business Understanding

A primeira etapa trata de entender o contexto da empresa, quais os desafios a serem resolvidos. É fundamental nessa etapa traçar os objetivos a serem alcançados e expectativas do cliente em relação à entrega. Defina aqui também quais as métricas de sucesso a serem alcançados. Conheça as pessoas envolvidas: saiba quem são as pessoas do time de tecnologia, quem é o responsável pela área de negócios e os tomadores de decisão.

Nessa etapa explore onde estão armazenados e organizados os dados, quais ferramentas a empresa utiliza e acessos que precisa para o projeto.



Data Understanding

A compreensão dos dados é uma das etapas que se gasta mais tempo por ter que coletar e organizar, iniciando realmente o trabalho de mineração dos dados. É importante o profissional validar a qualidade das informações obtidas e a viabilidade de se desenvolver o projeto.

É a fase de se fazer o estudo e levantamento dos conteúdos, conseguindo mapear se existem anomalias, quais os dados faltantes, como será a extração e qual o custo da análise. Uma das competências que ressalta é o lado investigativo de buscar respostas através das informações garimpadas e ter um olhar analítico de entender o que os dados representam.



Data Preparation

Depois do estudo e análise, chegou a hora de tratar esses dados brutos. Mas o que seria tratar os dados? É onde vai organizar quais os dados que serão modelados para o projeto, garantir que as informações estão alinhadas de acordo com o formato alinhado e o que será relevante para cruzar nas informações.

Aqui existem 4 etapas de preparação:

- **Data Selection:** a partir dos dados brutos, selecionar o que será modelado e documentar tudo;
- **Data Cleaning:** limpeza de fato, eliminando incoerências;
- **Construct Data:** com os dados finais obtidos na limpeza, vai identificar o que precisa construir e complementar para suas análises, criar campos ou tabelas, por exemplo;
- **Integrating Data:** fazer a integração de fontes distintas se for preciso.



Modeling

Pronto, agora é a hora de modelar!

Depois que fez a mineração inicial, limpou e tratou seus dados, o próximo passo é desenvolver o seu modelo.

Para isso, precisa decidir quais as técnicas de data mining que serão aplicadas, fazer testes em amostras e verificar se o modelo está com um bom desempenho computacional para seguir. Estes dados minerados serão os insumos para os algoritmos na predição do modelo do negócio.



Evaluation

Após criar modelos e algoritmos, nessa etapa fará a avaliação dos resultados e quais as variações podem surgir. É importante observar se o modelo criado atende aos objetivos iniciais traçados na etapa de entendimento do negócio.

É uma etapa que pode apresentar uma necessidade de modificar o modelo por não atender aos critérios estabelecidos, não tendo assim uma boa performance.



Deployment

Com o modelo desenvolvido, chegou a hora de colocá-lo em produção!

Na etapa de deployment, implementa-se o protótipo e o cliente faz o acompanhamento de sua performance após sua entrega.

Esse é o momento que a equipe entrega o modelo e organiza o conhecimento para poder ser repassado ao usuário.