

Data Science – Princípios e Técnicas

Março
2024



Data Science

Onde me encontrar:

<https://www.linkedin.com/in/marco-mialaret-junior/>

e

<https://github.com/MatmJr>

O Ciclo de Vida da Ciência de Dados

Data Science

O Ciclo de Vida da Ciência de Dados envolve várias etapas, desde a limpeza de dados até a modelagem e avaliação, para extrair insights e previsões úteis para objetivos comerciais.

As principais etapas são:

Data Science



Data Science

1. **Definição do problema:** O ciclo completo gira em torno do objetivo empresarial? O que você resolverá se não tiver um problema específico?

É extremamente importante compreender claramente o objetivo, pois esse será o objetivo final da análise.

Data Science

Após uma compreensão adequada, podemos definir o objetivo específico da análise que está alinhado com o objetivo geral do projeto.

Você precisa entender se o cliente deseja reduzir a perda de crédito, ou se ele quer prever o preço de uma mercadoria, etc.

Data Science

2. Coleta dos Dados: Após a compreensão do negócio, o próximo passo é a compreensão dos dados. Isso inclui uma coleta de todos os dados disponíveis.

Data Science

Aqui, você precisa trabalhar de perto com a equipe de negócios, pois eles estão certamente cientes de quais dados estão presentes, quais dados devem ser usados para esse problema comercial e outras informações.

Data Science

Este passo inclui descrever os dados, sua estrutura, relevância, tipo de dados. Explore os dados usando gráficos. Basicamente, extrair qualquer informação que você possa obter sobre os dados simplesmente explorando-os.

Data Science

3. **Pré-Processamento:** Em seguida, vem a etapa de preparação dos dados. Isso inclui etapas como escolher os dados aplicáveis, integrar os dados por meio da fusão dos conjuntos de dados, limpá-los, tratar os valores faltantes, tratar dados incorretos, também verificar a existência de outliers.

Data Science



Construir novos dados, derivar novos elementos a partir dos existentes. Formatar os dados para a estrutura desejada, eliminar colunas e características indesejadas.

A preparação dos dados é a etapa mais demorada e a mais importante em todo o ciclo de vida. Seu modelo será tão bom quanto seus dados.

Data Science

4. Análise Exploratória de Dados / Solução do Problema: Esta etapa inclui obter alguma ideia sobre a solução e os fatores que a afetam, antes de construir o modelo real. A distribuição dos dados dentro de variáveis diferentes de um indivíduo é explorada graficamente usando gráficos de barras.

Data Science



Relações entre diferentes aspectos são capturadas por meio de representações gráficas como gráficos de dispersão e mapas de calor. Muitas técnicas de visualização de dados são usadas extensivamente para explorar cada característica individualmente e combinando-as com outras características.

Data Science

5. Modelagem de Dados: A modelagem de dados é o coração da análise de dados. Um modelo recebe os dados organizados como entrada e fornece a saída desejada. Esta etapa inclui a seleção do tipo apropriado de modelo, seja um problema de classificação, regressão ou agrupamento.

Data Science

Após selecionar a família do modelo, dentre o número de algoritmos dentro dessa família, precisamos escolher cuidadosamente os algoritmos para implementar e executá-los. Precisamos ajustar os hiperparâmetros de cada modelo para alcançar o desempenho desejado.

Data Science

Também precisamos garantir que haja um equilíbrio adequado entre desempenho e generalização. Não desejamos que o modelo aprenda os dados e tenha um desempenho insatisfatório em novos dados.

Data Science

6. Avaliação do Modelo: Aqui, o modelo é avaliado para verificar se está pronto para ser implantado. O modelo é testado em dados não vistos, avaliado em um conjunto cuidadosamente pensado de métricas de avaliação.

Data Science

Também precisamos garantir que o modelo esteja em conformidade com a realidade. Se não obtivermos um resultado satisfatório na avaliação, devemos reiterar todo o processo de modelagem até alcançar o nível desejado de métricas.

Data Science

Assim como um humano, qualquer solução de ciência de dados, um modelo de aprendizado de máquina deve evoluir, ser capaz de melhorar com novos dados, adaptar-se a uma nova métrica de avaliação.

Data Science

Podemos construir vários modelos para um certo fenômeno, mas muitos deles podem ser imperfeitos. A avaliação do modelo nos ajuda a selecionar e construir um modelo ideal.

Data Science

7. Implantação do Modelo: O modelo, após uma avaliação rigorosa, é finalmente implantado no formato e canal preferidos. Este é o último passo no ciclo de vida da ciência de dados. Cada etapa no ciclo de vida da ciência de dados definida acima deve ser trabalhada cuidadosamente.

Data Science

Se qualquer etapa for executada de maneira inadequada, afetará a etapa subsequente e todo o esforço será desperdiçado. Por exemplo, se os dados não forem coletados corretamente, você perderá informações e não conseguirá construir um modelo ideal.

Data Science

Se os dados não forem limpos adequadamente, o modelo não funcionará. Se o modelo não for avaliado corretamente, falhará no mundo real. Desde a percepção do negócio até a implantação do modelo, cada etapa deve receber atenção, tempo e esforço adequados.

Dúvidas?



Telephone:

E-mail:

