CURSO: <BÁSICO EM MACHINE LEARNING>
Atividade 01 (ATIV-01)
Hosana Fernandes Gomes

1. Explique, com suas palavras, o que é machine learning?
Sub-ramo da inteligência artificial que utiliza algoritmos para identificar e classificar padrões e para estabelecer predições com base em dados e em eventos anteriores. Os modelos de ML são retroalimentáveis, isto é, capazes de sofrerem melhoramentos à medida que se aumenta o input de dados e o repositório de "experiências" anteriores .

2. Explique o conceito de conjunto de treinamento, conjunto de validação e conjunto de teste em machine learning.
São subconjuntos de um dataset a ser inserido em um algoritmo de ML. O conjunto de treinamento deve ser o mais robusto possível, pois deve representar a complexidade dos eventos reais. Ainda, a partir desse conjunto serão reajustados os parâmetros do modelo para minimizar os erros e as perdas nas próximas etapas.

Em seguida, o conjunto de validação serve para testar o seu modelo diante de dados ainda desconhecidos a fim de ajustar os hiperparâmetros, os responsáveis por melhorar a precisão do seu modelo, evitando *under* ou *overfitting*. Por fim, o conjunto de teste é utilizado estritamente após as etapas anteriores, fazendo uso de dados novos e independentes apenas para avaliar o desempenho do modelo e utilizar esses resultados para estabelecer comparações com outras propostas que também utilizam ML.

3. Explique como você lidaria com dados ausentes em um conjunto de dados de treinamento.
Depende do tipo de problema que estou lidando. Geralmente, os NaN são substituídos por 0, pela moda ou mediana, mas a substituição deve ser feita de acordo com o contexto, de modo a buscar a melhor maneira de representar os eventos da vida real. Entretanto, se a quantidade de valores ausentes for insignificante em relação ao total de dados, podemos excluí-los do conjunto de treinamento.

4. O que é uma matriz de confusão e como ela é usada para avaliar o desempenho de um modelo preditivo?
É uma tabela inserida no modelo para testar a sua precisão, acurácia e desempenho (f-score e recall). Através desses resultados, é possível avaliar a qualidade do modelo e implementar melhorias caso necessário.

5. Em quais áreas (tais como construção civil, agricultura, saúde, manufatura, entre outras) você acha mais interessante aplicar algoritmos de ML?
Algoritmos de ML podem ser úteis no urbanismo para análise de indicadores e predição de cenários com base na implementação de políticas públicas. Na área da saúde, podem ser úteis para estabelecer diagnósticos de imagem com base em resultados anteriores, alimentados no modelo nas etapas de treinamento.