**Relatório <010> - < 10 - Prática: Lidando com Dados do Mundo Real (II) >**

<Kawan Machado>

# Descrição da atividade

## More Data Mining and Machine Learning Techniques

### K-Nearest-Neighbors Concepts

Iniciando por esse conceito, fica evidente como as técnicas de aprendizado de máquina podem ser tão simples. A ideia é classificar os pontos dos dados baseado na distância entre si. Encontre os vizinhos mais próximos e agrupe os baseado em suas distânicas.

FIGURA 1 - K-NEAREST-NEIGHBORS

### Activity Using KNN to predict a rating for a movie

### Dimensionality Reduction; Principal Component Analysis (PCA)

### Activity PCA Example with the Iris data set

### Data Warehousing Overview ETL and ELT

### Pac-Man Example.html

### Python Markov Decision Process Toolbox.html

### Cat and Mouse Example.html

### Reinforcement Learning

### Activity Reinforcement Learning & Q-Learning with Gym

### Understanding a Confusion Matrix

### Measuring Classifiers (Precision, Recall, F1, ROC, AUC)

## Dealing with Real-World Data

### Measuring Classifiers (Precision, Recall, F1, ROC, AUC)

### Activity K-Fold Cross-Validation to avoid overfitting

### Data Cleaning and Normalization

### Activity Cleaning web log data

### Normalizing numerical data

### Activity Detecting outliers

### Feature Engineering and the Curse of Dimensionality

### Imputation Techniques for Missing Data

### Handling Unbalanced Data Oversampling, Undersampling, and SMOTE

### Binning, Transforming, Encoding, Scaling, and Shuffling

# Conclusões

# Referencias

As referências foram apenas os vídeos da atividade.