

# Nanodegree Engenheiro de Machine Learning

## Proposta de projeto final

---

Gleber Baptistella

Fevereiro de 2018

## Proposta

---

### Histórico do assunto

A área de Reconhecimento de Atividades Humanas (RAH) está em franca expansão, com dispositivos e sensores cada vez mais fazendo parte do dia-a-dia das pessoas.

Suas pesquisas podem ser aplicadas em diversas áreas, entre elas da saúde (melhorando a detecção e diagnósticos de doenças) e de práticas esportivas, visando a melhoria do desempenho do indivíduo.

### Descrição do problema

Este projeto tem como objetivo identificar uma determinada postura de um indivíduo após a coleta de dados através de sensores que foram previamente dispostos no corpo da pessoa. O embasamento deste projeto encontra-se em <http://groupware.les.inf.puc-rio.br/har>, bem como o dataset utilizado.

Irei avaliar outros algoritmos diferentes do usado no artigo da PUC Rio, buscando um desempenho superior ao obtido.

### Conjuntos de dados e entradas

O conjunto de dados selecionado encontra-se disponível em <http://groupware.les.inf.puc-rio.br/static/har/dataset-har-PUC-Rio-ugolino.zip> e está descritos abaixo.

Dados Pessoais:

- user: nome da pessoa
- gender: sexo da pessoa
- how\_tall\_in\_meters: altura da pessoa
- weight: peso da pessoa
- body\_mass\_index: Índice de massa corpórea da pessoa

Dados dos sensores:

São colocados 4 sensores no corpo do indivíduo.

Para cada sensor são definidos os eixos x, y e z e no dataset temos para o sensor 1 os dados, x1, y1 e z1. Para o sensor 2, os dados x2, y2 e z2, e assim por diante. A posição de cada sensor é:

- Sensor 1: cintura
- Sensor 2: coxa esquerda
- Sensor 3: canela direita
- Sensor 4: braço direito.

Por fim, temos a feature "class". Trata-se da variável que queremos prever e que pode assumir os seguintes valores: 'sitting', 'sittingdown', 'standing', 'standingup', 'walking'

## Descrição da solução

A solução encontrada para o problema original está descrita neste [link](#). Foi utilizada uma árvore de decisão C4.5 com o ensemble method AdaBoost.

Neste projeto iremos testar outros algoritmos, especificamente o K-Nearest Neighbors (KNN), Random Forest e Gaussian naive bayes para compararmos com os resultados originais.

## Modelo de referência (benchmark)

Neste projeto teremos dois modelos de referência:

1. O resultado do artigo original
2. A execução sem refinamento dos outros três algoritmos.

Enquanto tentamos superar o resultado do artigo original (o que pode não ser possível), tentaremos refinar os outros três algoritmos através de mudanças em seus hiper-parâmetros para obtermos resultados melhores.

## Métricas de avaliação

As métricas para avaliação dos modelos serão:

1. Recall
2. Precision
3. F1
4. ROC

## Design do projeto

Tendo definido o problema e o dataset como exposto acima, segue minha proposta de fluxo de trabalho.

Em primeiro lugar irei fazer a exploração dos dados em busca missing values ou valores que não estão condizentes com o restante dos dados. Esta fase acredito ser rápida já que os dados coletados dos sensores e disponibilizados a partir do artigo original já devem estar consistentes.

Em seguida, irei trabalhar os dados para deixá-los no formato apropriado para a aplicação dos algoritmos.

Todas as features que são string e serão utilizadas nos algoritmos serão convertidas para valores numéricos mais especificamente as variáveis gender e class. A variável user será excluída do dataset.

Após o tratamento das variáveis, irei executar os algoritmos propostos em sua forma padrão, sem nenhuma parametrização adicional.

O resultado desta execução será a baseline para o restante do projeto.

Com a baseline definida, irei começar um ciclo de feature engineering, parametrização e execução dos algoritmos e análise de resultados

Por fim, na última parte do projeto consiste na apresentação dos resultados dos ciclos e comparação de resultados com o artigo original.

## Referências

Ugulino, W.; Cardador, D.; Vega, K.; Velloso, E.; Milidui, R.; Fuks, H. [Wearable Computing: Accelerometers' Data Classification of Body Postures and Movements](#). Proceedings of 21st Brazilian Symposium on Artificial Intelligence. Advances in Artificial Intelligence - SBIA 2012. In: Lecture Notes in Computer Science. , pp. 52-61. Curitiba, PR: Springer Berlin / Heidelberg, 2012. ISBN 978-3-642-34458-9. DOI: 10.1007/978-3-642-34459-6\_6.