

# Desafio AGE – Bioengenharia

## Soft Biometrics System for User Identification using Smartphone Gait Data

**Costa&Rocha:**

João António Fernandes da Costa, [bio12046@fe.up.pt](mailto:bio12046@fe.up.pt)

João Salgueiro Rocha, [bio10009@fe.up.pt](mailto:bio10009@fe.up.pt)

- **Introdução**

- Descrição do Desafio
- Descrição do Dataset

- **Pré-processamento**

- Interpolação Linear

- **Deteção de caminhada**

- STFT
- Clustering com K-means

- **Extração de Features**

- Windowing
- Features estatísticas
- MFCCs

- **Seleção de Features**

- Mutual Info Score

- **Classificação**

- SVM
- Cross-Validation e Hyperparameter Tuning

- **Resultados**

- Train Set
- Test Set

- **Conclusões**

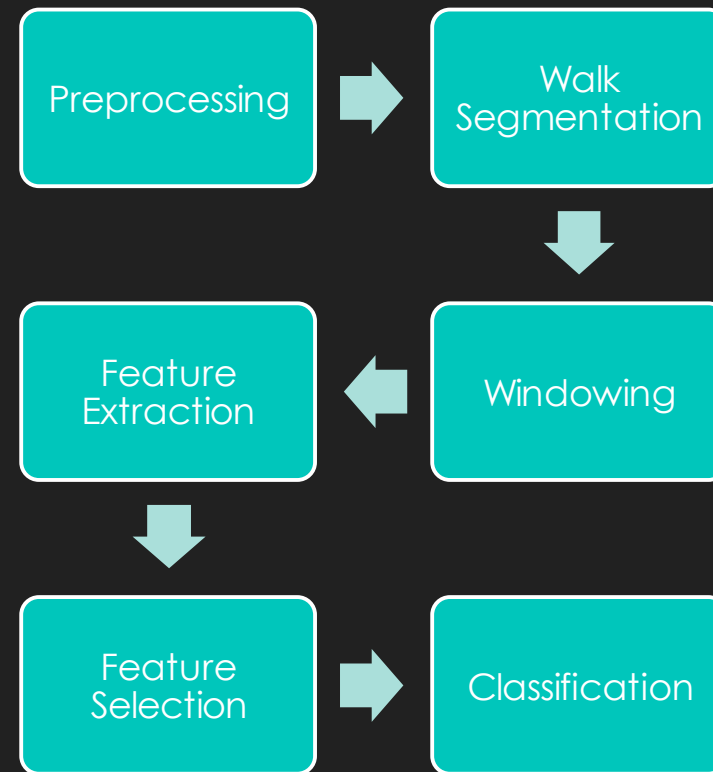
# Introdução

Descrição do Desafio

Descrição do Dataset

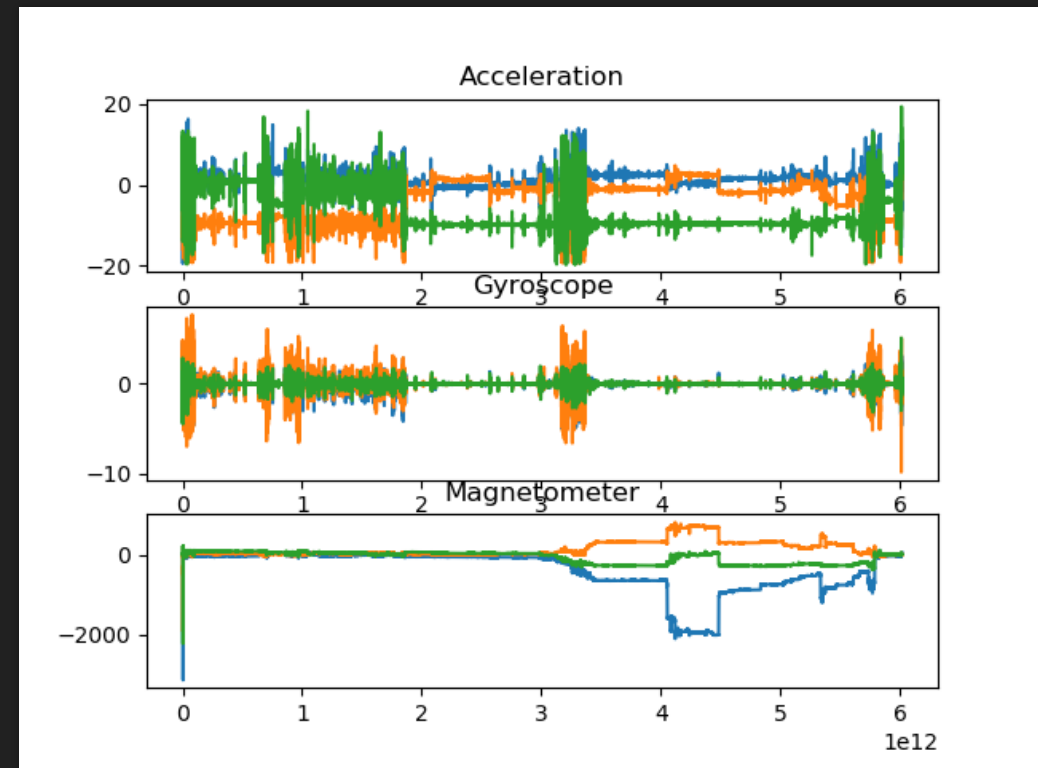
# Descrição do Desafio

- Identificação de utilizadores através de dados de smartphone
  - Acelerómetro, Giroscópio e Magnetómetro
- Identificação de períodos de marcha
- Extração de características de passada
- Classificação



# Descrição do Dataset

- 26 utilizadores
- Número de sessões variável
- Duração das sessões variável
- Utilização natural do smartphone
- Dataset de teste:
  - 10 sessões de um só utilizador



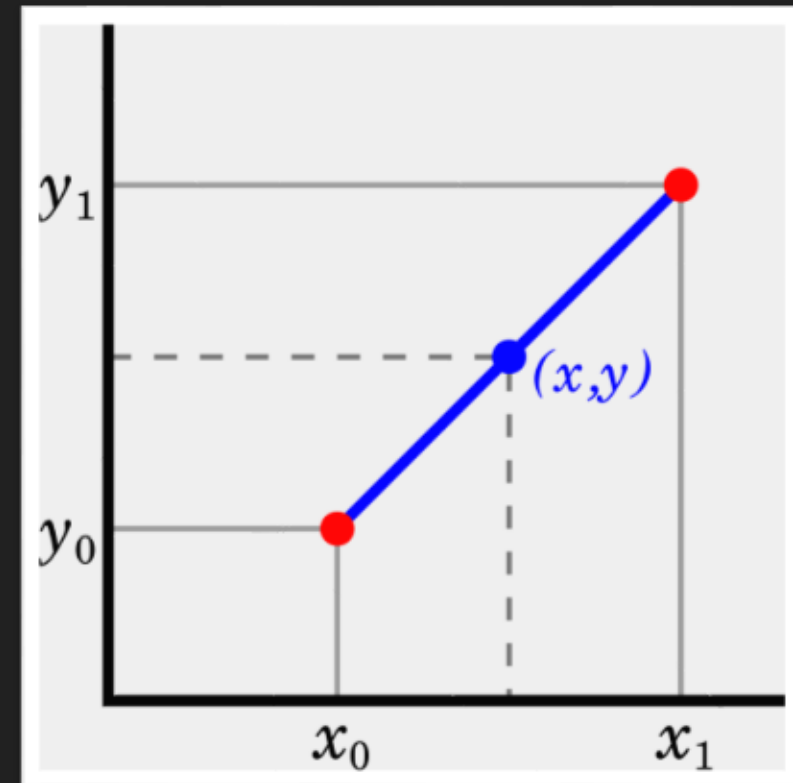
Leituras dos sensores para uma sessão

# Pré-Processamento

Interpolação Linear

# Interpolação Linear

- Normalização da frequência de amostragem a 100 Hz para todos os sensores
  - Igual nº de amostras para todos os sensores
  - Elimina flutuações de amostragem devido a imperfeições dos smartphones



Interpolação Linear

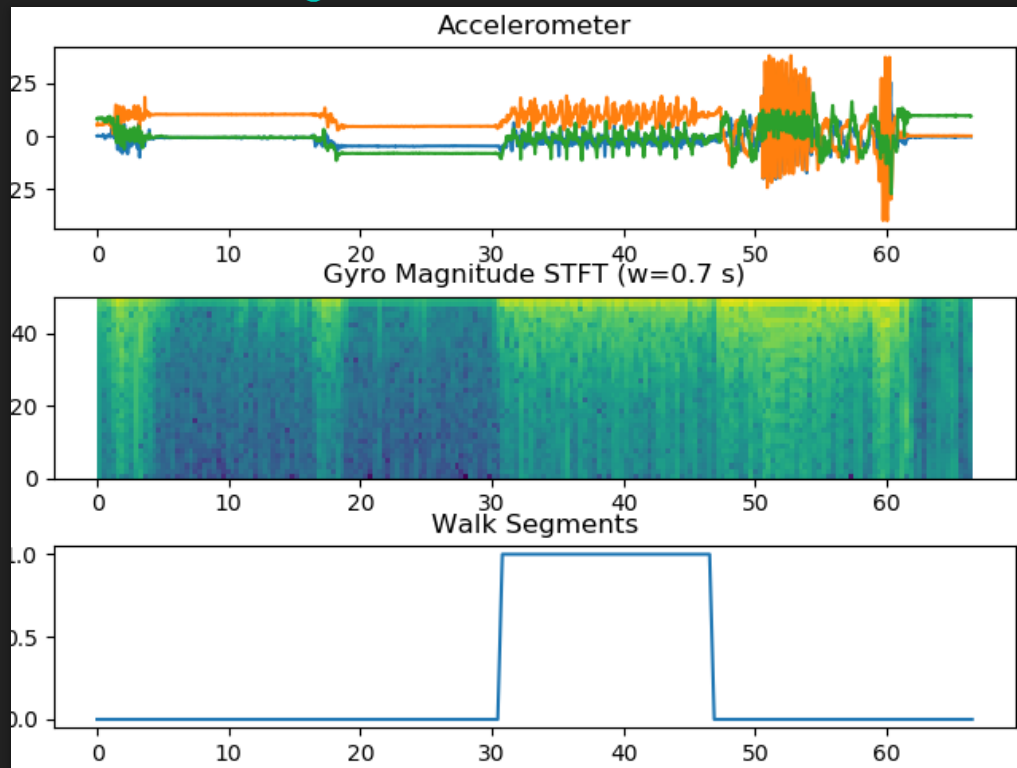
# Deteção de Caminhada

STFT – Short-Time Fourier Transform

Clustering com K-means



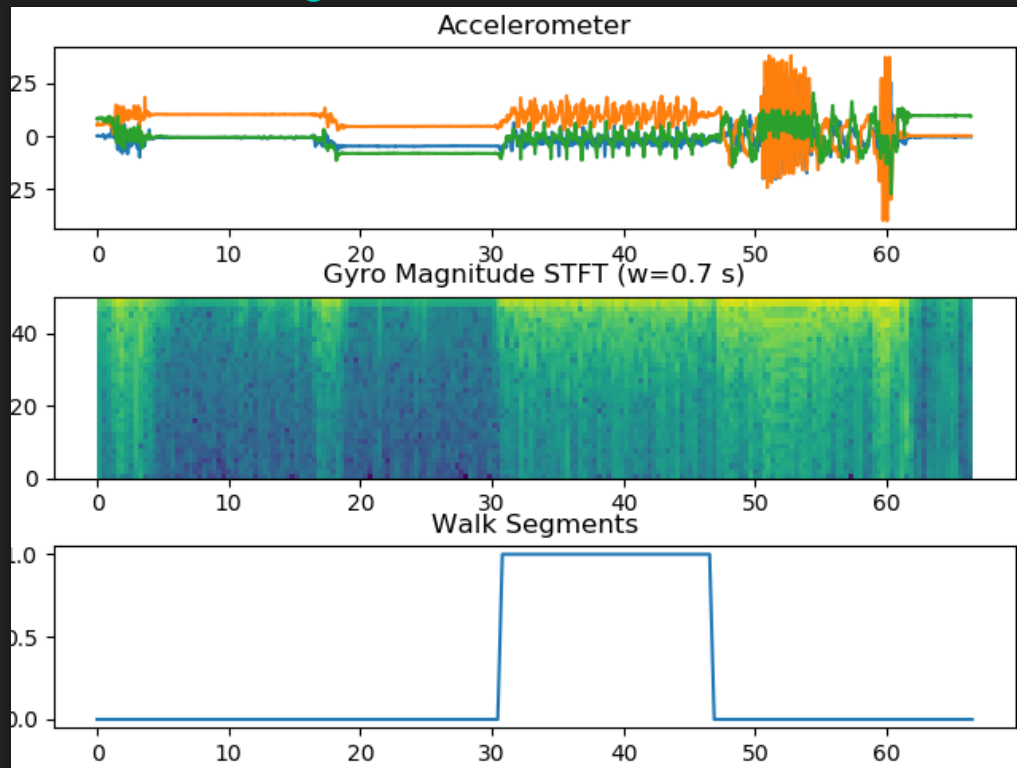
# STFT – Short-Time Fourier Transform



Exemplo de segmentação de caminhada

- Calculada a STFT da magnitude do giroscópio ao longo do tempo
- Espectro de frequências dependente da atividade realizada
  - Utilizador parado – baixa energia em todas as frequências
  - Utilizador a andar – energia mais concentrada nas baixas frequências (<7Hz)

# Clustering com K-means



Exemplo de segmentação de caminhada

- Coeficientes da STFT usados como features para clustering com K-means (K=5)
- Identificados clusters com perfis de energia semelhantes a passada
- Segmentação dos sinais dos sensores

# Extração de Features

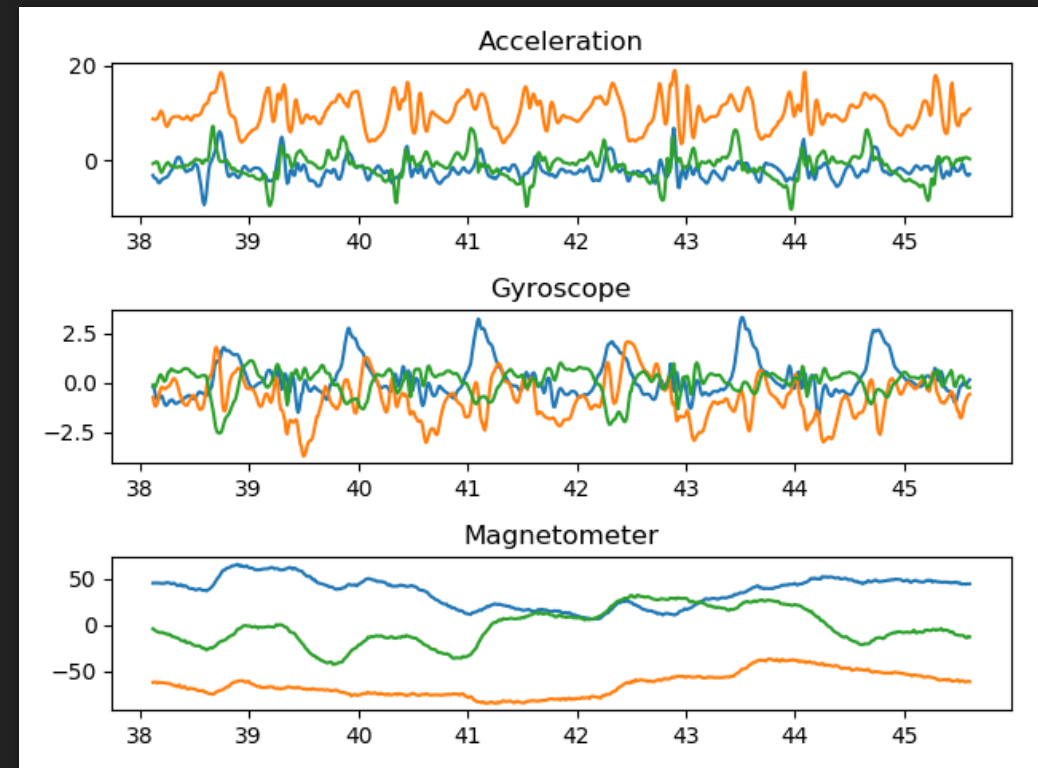
Windowing

Features Estatísticas

MFCCs

# Windowing

- Segmentos de caminhada são cortados em janelas sucessivas
  - 7,5 s de duração
  - 50% de overlap entre janelas



Janela de caminhada

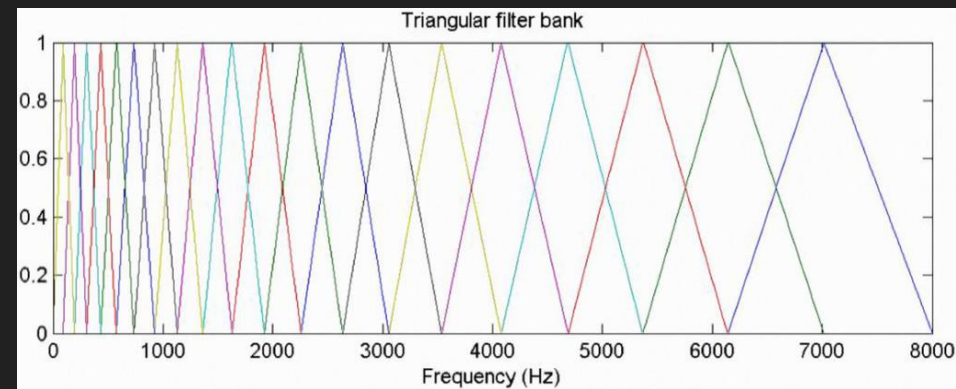
# Features

## Estatísticas

- Features extraídas dos eixos x, y, z e magnitude do acelerómetro e giroscópio (136 features)
  - Média
  - Máximo, Mínimo e Diferença
  - Desvio-padrão
  - RMS
  - Zero-crossings
  - Histograma (bins = 10)

## MFCCs

- Espectro de MFCCs extraído para cada eixo x,y,z e magnitude do acelerómetro e giroscópio (518 features)
  - Processamento de Voz
  - Janelas de 1,5s, sem overlap



Filtros triangulares para MFCC

# Seleção de Features

Feature Scaling  
Mutual Info Score

# Feature Scaling

## StandardScaler

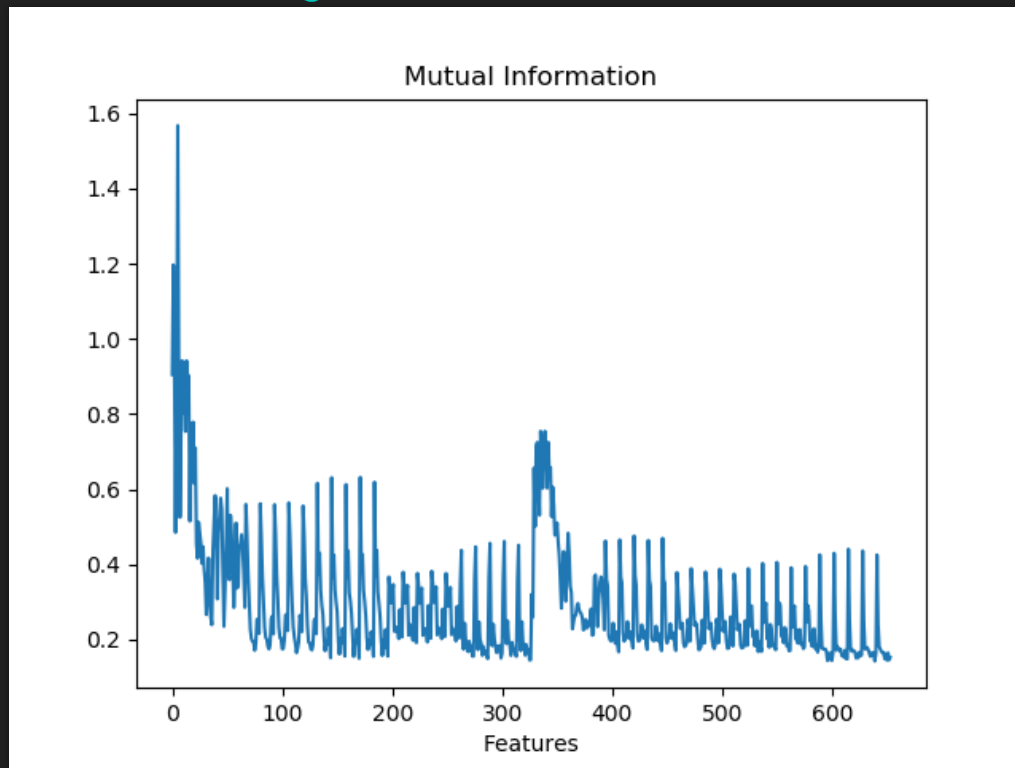
- Normaliza cada feature para ter média 0 e desvio-padrão 1
  - Frequentemente usada
  - Outliers podem criar enviesamento na média

## RobustScaler

- Normaliza cada feature de acordo com a sua mediana e amplitude interquartil
  - Mais robusta a outliers
- Pré-processamento escolhido
  - Alta variabilidade nas features da passada para um mesmo utilizador, devido a outliers



# Mutual Information



Mutual Information das features usadas

- Quantifica informação que se pode obter sobre uma variável a partir de outra
  - Quantificar informação útil que se pode obter de uma feature para o problema de classificação
- Features ordenadas segundo Mutual Info
- Selecionam-se as 45% melhores



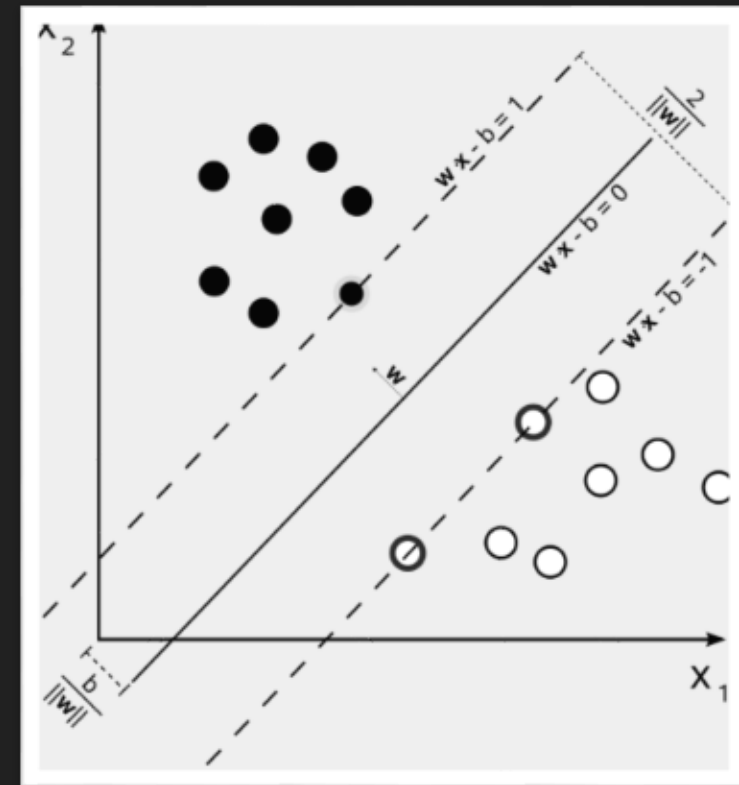
# Classificação

SVM

Cross-Validation e Hyperparameter Tuning

# SVM

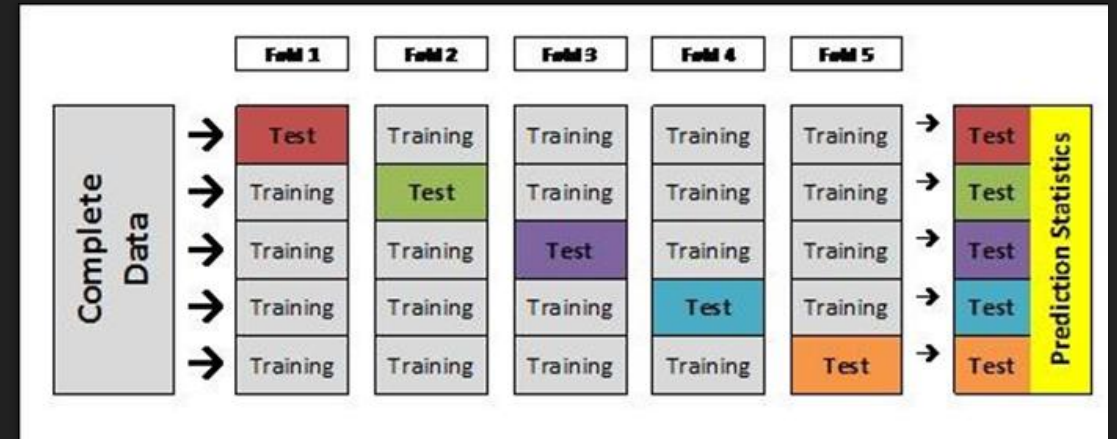
- Classificadores testados:
  - Naïve Bayes
  - SVM
- One Vs Rest SVM
- SVM mostrou ser o melhor classificador



Support Vector Machine

# Cross-Validation e Hyperparameter Tuning

- Para validar o classificador e ajustar os hiperparâmetros, foi usada Grid Search Cross Validation
  - 3 folds
  - Ajuste da percentagem de features a selecionar pela sua Mutual Information
  - Ajuste de C e gamma da SVM
- Melhores parâmetros:
  - Percentagem = 45%
  - C = 10
  - gamma = 0.001



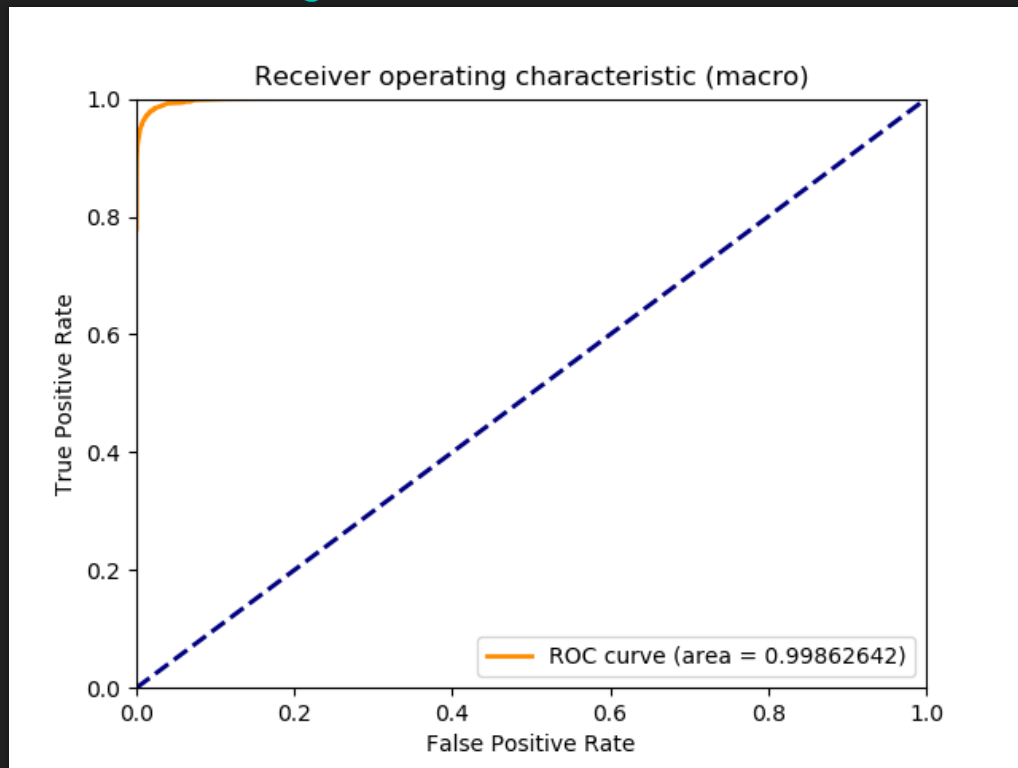
Cross Validation, com 5 folds

# Resultados

Train Set

Test Set

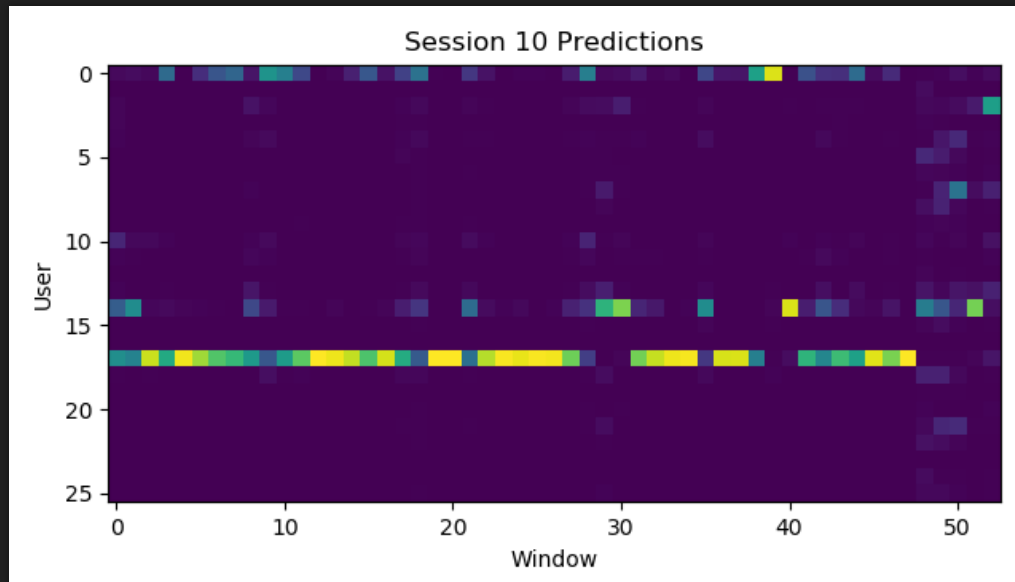
# Train Set



	precision	recall	f1-score	support
1	0.98	0.89	0.93	46
2	0.90	0.88	0.89	78
3	0.94	0.98	0.96	221
4	0.99	0.99	0.99	756
5	0.99	0.96	0.97	312
6	0.81	0.82	0.82	89
7	0.91	0.93	0.92	103
8	1.00	0.98	0.99	290
9	0.89	0.97	0.93	430
10	0.95	0.97	0.96	303
11	0.98	0.95	0.96	135
12	1.00	0.86	0.92	21
13	0.95	0.83	0.89	24
14	0.97	0.88	0.92	109
15	0.86	0.89	0.87	245
16	1.00	1.00	1.00	1101
17	1.00	1.00	1.00	206
18	0.99	0.95	0.97	115
19	0.99	0.94	0.97	352
20	0.94	0.88	0.91	67
21	0.94	0.93	0.93	97
22	0.95	0.95	0.95	270
23	0.67	0.74	0.70	86
24	0.99	1.00	0.99	492
25	0.78	0.68	0.73	57
26	0.99	0.97	0.98	137
avg / total	0.96	0.96	0.96	6142

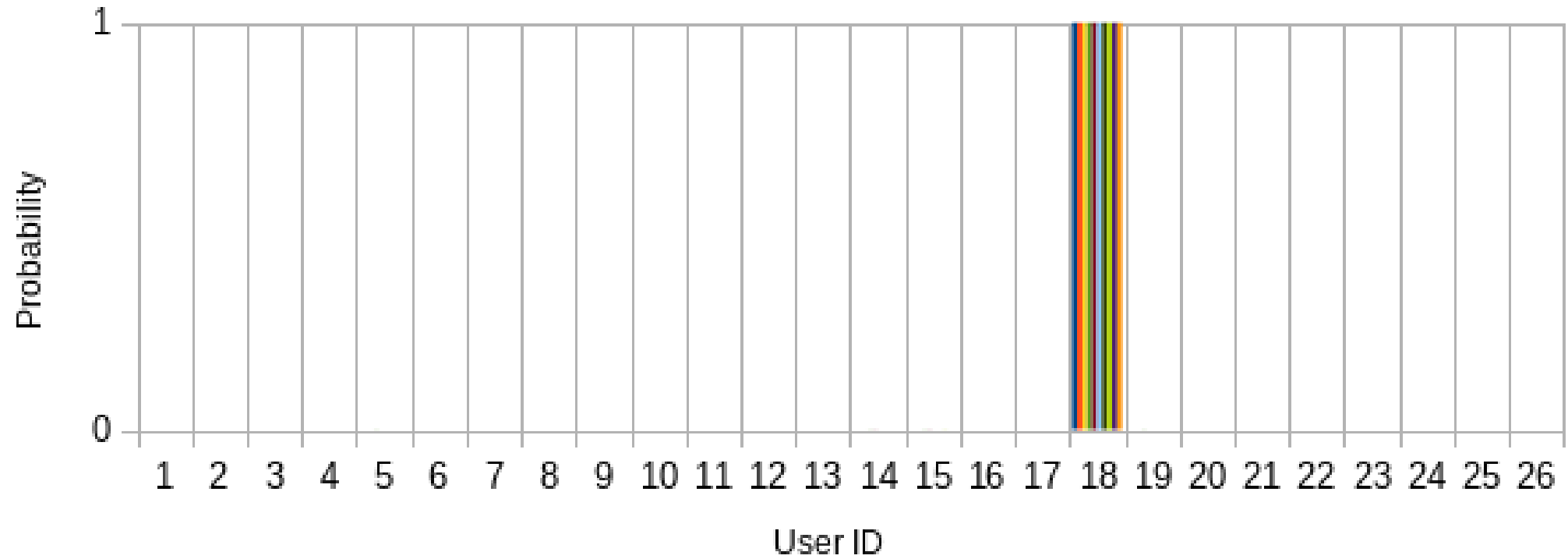
Classification Report  
(accuracy = 95.9%)

# Test Set



- Cada sessão é sujeita à mesma pipeline
- Previsões por janela
- Probabilidades de cada utilizador são multiplicadas e normalizadas.

Predictions for Test Set



■ Probability\_Folder1 ■ Probability\_Folder2 ■ Probability\_Folder3 ■ Probability\_Folder4  
■ Probability\_Folder5 ■ Probability\_Folder6 ■ Probability\_Folder7 ■ Probability\_Folder8  
■ Probability\_Folder9 ■ Probability\_Folder10

# Conclusões



# Conclusões

- Features mais importantes são as do eixo x dos sensores
- Acelerómetro apresenta mais informação
- Features estatísticas simples, como a média, já são suficientes para classificação
- Segmentação de períodos de marcha pode ser melhorado
  - Problema supervisionado p.e.

# Bibliografia

- Brajdic, Agata, and Robert Harle. "Walk detection and step counting on unconstrained smartphones." Proceedings of the 2013 ACM international joint conference on Pervasive and ubiquitous computing. ACM, 2013.
- Nickel, Claudia, Tobias Wirtl, and Christoph Busch. "Authentication of smartphone users based on the way they walk using k-nn algorithm." Intelligent Information Hiding and Multimedia Signal Processing (IIH-MSP), 2012 Eighth International Conference on. IEEE, 2012.
- Derawi, Mohammad Omar, et al. "Unobtrusive user-authentication on mobile phones using biometric gait recognition." Intelligent Information Hiding and Multimedia Signal Processing (IIH-MSP), 2010 Sixth International Conference on. IEEE, 2010.

Obrigado Pela Atenção!