

SERVIÇO PÚBLICO FEDERAL  
MINISTÉRIO DA EDUCAÇÃO  
UNIVERSIDADE FEDERAL FLUMINENSE  
INSTITUTO DE CIÊNCIAS EXATAS  
COORDENAÇÃO DO CURSO DE FÍSICA

# **DEFESA DE TRABALHO DE CONCLUSÃO DE CURSO**

*Autor: Vinícius César Peres Silva Ferreira*

## **“TÉCNICAS DE APRENDIZADO DE MÁQUINA APLICADAS AOS DADOS DO PROJETO ADACA”**

*Orientadora: Profa. Dra. Vera Lúcia Prudência dos Santos Caminha*

### **Banca Examinadora:**

<i>Vera Lúcia Prudência dos Santos Caminha, D.C.</i>	<i>(Orientadora-UFF)</i>
<i>Adriano de Oliveira Caminha, D.C.</i>	<i>(UFF)</i>
<i>Marina Sequeiros Dias de Freitas, D.C.</i>	<i>(UFF)</i>

### **Resumo:**

Este trabalho visa aplicar os modelos de Floresta Aleatória, Árvore de Decisão e Regressão Logística na análise dos dados obtidos pelo projeto de pesquisa e extensão ADACA (Ambiente Digital de Aprendizagem para Crianças Autistas), analisando características construídas a partir das posições do mouse ao longo das sessões, como séries temporais da velocidade do cursor e estatísticas descritivas dessas séries e do seu espaço recíproco obtidos pela Transformada Discreta de Fourier, comparando os seus desempenhos e escolhendo o modelo que melhor realiza a tarefa de classificar o conjunto de dados através de categorias rotuladas. Antes, os dados são explorados e possuem sua dimensionalidade reduzida pela técnica do PCA (Análise dos Componentes Principais) e são agrupados pelo algoritmo das K-médias. Com isso, pode-se verificar a eficácia das técnicas de Aprendizado de Máquina para análise dos dados do projeto ADACA. Os dados foram agrupados em 2 clusters pelo algoritmo das K-médias e analisados pela suas projeções sobre os dois primeiros componentes principais. Foi escolhido para estudo o cluster com mais dados. Foram construídas 23 estatísticas descritivas das Séries Temporais da velocidades do cursor, incluindo de seu espaço recíproco, sendo realizada uma redução de dimensionalidade via PCA para 10 Componentes Principais de modo a manterem mais de 80 % da variância total dos dados. Essas características alimentaram os modelos acima, validados pela AUC ROC e pela pontuação F1. O modelo mais estável na questão de generalização e com excelente desempenho foi o da Regressão Logística, enquanto os demais modelos, embora eficientes e com bom desempenho, padecem de alto sobreajuste sobre os dados de treinamento.

***Dia: 6ª feira, 05/12/2025***

***Hora: 14h00 (Brasília - GMT-3)***

***Sala: 206 C***