

# Universidade de Pernambuco

## Programa de Pós-Graduação em Engenharia da Computação (PPGEC)

### Proposta de Tese de Doutorado

**Área: Computação Inteligente**

**Título: Desenvolvendo uma Ferramenta Baseada em Machine Learning para Recuperação de Arquivos JPEG e PNG Fragmentados Após Formatação**

**Orientador(a): Sidney Marlon Lopes de Lima (sidney.lima@ufpe.br)**

#### Descrição:

Os sistemas de armazenamento de dados dividem arquivos em clusters que seguem uma sequência lógica. Conforme os arquivos são adicionados, formatados ou excluídos de um disco, ocorre a fragmentação externa dos dados iniciais, gerando espaços livres para novos arquivos. Este fenômeno de fragmentação é especialmente relevante na computação forense, dada a crescente quantidade de arquivos armazenados digitalmente pelos usuários [1].

No contexto dos discos de armazenamento, a fragmentação ocorre quando arquivos são repetidamente excluídos ou formatados, resultando em espaços vazios entre blocos de dados sequenciais. Essa fragmentação distribui fragmentos de arquivos pelo sistema de armazenamento, tornando-os indistinguíveis na ausência de metadados adequados [1] [2].

Usando algoritmos de aprendizado de máquina, como Máquinas de Vetores de Suporte (SVM), a ferramenta será projetada para identificar e reconstruir fragmentos de arquivos com acurácia, mesmo na ausência de metadados completos [3]. Este projeto contribuirá significativamente para o campo da computação forense, ajudando a resolver crimes digitais que envolvem arquivos fragmentados [4]. A pesquisa abordará a análise da fragmentação de arquivos, o desenvolvimento de algoritmos de recuperação, a implementação de uma interface amigável para usuários finais e a validação da ferramenta em diversos cenários de fragmentação e formatação [5]. Além disso, apresenta a proposta inicial da ferramenta que é recuperar dados formatados.

#### Referências Bibliográficas:

- [1] SENCAR, Husrev T.; MEMON, Nasir. Identification and recovery of JPEG files with missing fragments. *digital investigation*, v. 6, p. S88-S98, 2009.
- [2] GARFINKEL, Simson L. Carving contiguous and fragmented files with fast object validation. *digital investigation*, v. 4, p. 2-12, 2007.
- [3] MOCK, Ken C.; MALINOSKI, Steven M. "File carving: Identifying file types and fragments from binary data." *Digital Investigation*, vol. 4, pp. 60-67, 2007
- [4] CHANG, Po-Chi; CHEN, Chung-Wei; LIU, Chien-Chao. "Improved carving of fragmented files with sequential hypothesis tests." *ACM Transactions on Information Systems Security*, vol. 17, no. 1, pp. 1-27, 2014.
- [5] MOSS, Brett; CHEN, Kuang-Yu; DUNPHY, Paul; CLAYTON, Richard; ANDERSON, Ross. "Exploring practical issues in JPEG forensics." In *IEEE International Conference on Multimedia and Expo Workshops (ICMEW)*, pp. 1-6, 2012.