



**Universidade do Minho**

Departamento de Informática

Mestrado [Integrado] em Engenharia Informática

Perfil de Sistemas Inteligentes

Aprendizagem e Extração de Conhecimento

4º/1º Ano, 1º Semestre

Ano letivo 2018/2019

Ficha Prática nº 3

12 de novembro de 2018

**Tema** Data Preprocessing

**Enunciado** Nesta ficha pretende-se que o aluno adquira conhecimentos associados à preparação e gestão de dados, como forma de melhorar a performance de treino de modelos de machine learning em linguagem Python.

**Tarefas** Utilize o respectivo dataset do enunciado para a resolução desta Ficha Prática.

Ao analisar o dataset, e atendendo ao objectivo do projecto (capacidade de prever o nível de stress dos estudantes – PSS - durante exames de alto risco), verificamos que o conjunto de dados apresentam algumas limitações (variações de grande escala entre as diferentes features, dados em falta, apresentação de features com baixa variação ou pouco relevantes, etc.).

Ex.1) Treine um modelo de SVM e verifique a performance desse modelo (utilize o método train/test 80%-20%)

Ex.2) Utilize o seguinte conjunto de técnicas leccionadas na aula, como forma de preparar o dataset:

- Missing Value Replacement
- Normalization
- Variance Threshold
- Feature Selection - Recursive Feature Elimination

Ex.3) Treine um novo modelo de SVM com o dataset preparado no ex.2 e verifique a performance desse modelo (utilize o método train/test 80%-20%)