

Programa de Disciplina Especialização em Ciência dos Dados

Módulo: III

Disciplina: Técnicas de Redução de Dimensão

Carga Horária: 15 horas (9h teóricas, 4h práticas e 2h de avaliação)

Ofertante: Departamento de Engenharia de Produção – DEENP/UFOP

Objetivo:

Capacitar o estudante na preparação de bases de dados e na eliminação de informação estatística redundante de um conjunto de dados focando na interpretação dos dados e na construção de modelos de maior acurácia.

Ementa:

Preparação de base de dados. Aplicabilidade das técnicas de redução de dimensão. Métodos de seleção e extração de indicadores. Análise Discriminante Linear (LDA). Análise de Componentes Principais (PCA). Métodos não lineares. Implementação em Python. Implementação em R Aplicações na indústria siderúrgica.

Conteúdo Programático:

1. Preparação de base de dados.
2. Justificativa de uso das TRD.
3. Técnicas de Redução de Dimensão:
 - a) Seleção de indicadores.
 - b) Extração de indicadores.
 - c) Métodos supervisionados.
 - d) Métodos não supervisionados.
 - e) Métodos lineares.
 - f) Métodos não-lineares.
4. Análise de Componentes Principais:
 - a) Embasamento matemático.
 - b) Variância total e explicada.

- c) Visualização.
 - d) Análise Fatorial e Escalonamento Multidimensional.
5. Análise Discriminante Linear:
- a) Embasamento matemático.
 - b) Variância total e explicada.
 - c) Visualização.
 - d) Outros usos da LDA.
6. Aplicações na indústria siderúrgica:
- a) Caso I - Identificação da sazonalidade da variabilidade não-intencional do fluxo de informação de fornecedores.
 - b) Caso II - Melhoria na qualidade dos aços.
 - c) Caso III - Classificação da qualidade da soldagem.
7. TRD em Python.
8. TRD em R.

Metodologia:

A parte teórica desta disciplina terá a finalidade do embasamento matemático formal e o posicionamento das ferramentas dentro do conjunto de técnicas de aprendizado de máquina. A parte prática buscará capacitar o aluno para utilizar as bibliotecas existentes e ser capaz de analisar dados através dos métodos PCA e LDA. Serão apresentados também estudo de casos reais sobre a aplicação destas técnicas em problemas da indústria siderúrgica.

Atividade Prática Proposta:

Os alunos, em grupo, deverão analisar os dados simulados relativos a um processo de laminação e, a partir das ferramentas estudadas, identificar as componentes de maior importância no processo de formação de defeitos.

Softwares:

1. Os métodos estudados serão implementados utilizando o Python e o software R.

Bibliografia:

ALPAYDIN, E. *Introduction to Machine Learning*. 3. ed. EUA: MIT Press (MA), 2014. 613 p. ISBN 978-026202-818-9.

BISIO, G.; RUBATTO, G. Process improvements in iron and steel industry by analysis of heat and mass transfer. *Energy Conversion and Management*, v. 43, n. 2, p. 205–220, 2002. ISSN 0196-8904. Disponível em: <<http://www.sciencedirect.com/science/article/pii/S0196890401000267>>. Acesso em: 16 abr. 2020.

IZENMAN, A. J. *Modern Multivariate Statistical Techniques: Regression, Classification, and Manifold Learning*. 2. ed. New York: Springer, 2013. 731 p. ISBN 978-038778-188-4.

NAKAGAWA, Y. et al. Quality improvement of steel products by using multivariate data analysis. In: *SICE Annual Conference 2007*. IEEE, 2007. p. 2428–2432. Disponível em: <<https://ieeexplore.ieee.org/document/4421396>>. Acesso em: 16 abr. 2020.

PEDREGOSA, F. et al. Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research*, v. 12, n. 85, p. 2825–2830, out. 2011. Disponível em: <<http://jmlr.org/papers/v12/pedregosa11a.html>>. Acesso em: 15 abr. 2020.

PRATS-MONTALBÁN, J. M. et al. A comparison of different discriminant analysis techniques in a steel industry welding process. *Chemometrics and Intelligent Laboratory Systems*, v. 80, n. 1, p. 109–119, 2006. ISSN 0169-7439. Disponível em: <<http://www.sciencedirect.com/science/article/pii/S016974390500119X>>. Acesso em: 15 abr. 2020.

RASCHKA, S.; MIRJALILI, V. *Python machine learning: Machine Learning and Deeping Learning with Python, scikit-learn, and TensorFlow*. 2. ed. Birmingham: Packt Publishing, 2017. 622 p. ISBN 978-178712-593-3.

THORNHILL, N. F.; NAIM, M. M. An exploratory study to identify rogue seasonality in a steel company's supply network using spectral principal component analysis. *European Journal of Operational Research*, v. 172, n. 1, p. 146–162, 2006. ISSN 0377-2217. Disponível em: <<http://www.sciencedirect.com/science/article/pii/S0377221704007817>>. Acesso em: 15 abr. 2020.