

ANÁLISE E VISUALIZAÇÃO DA DISTRIBUIÇÃO GAUSSIANOS COM A BIBLIOTECA PANDAS PARA SELEÇÃO DE ATRIBUTOS

Eduardo Matheus Rodrigues de Oliveira¹, Renan Caldeira Menechelli¹, Elvio Gilberto da Silva¹, Rosane Maria Lima Araújo¹

¹Centro de Ciências Exatas e Sociais Aplicadas – Universidade do Sagrado Coração – eduardomr.oliveira@gmail.com; renan.menechelli@gmail.com; egilberto@uol.com.br; rosanemlaraujo@terra.com.br

Tipo de pesquisa: TCC.

Agência de Fomento: Não há

Área do conhecimento: Exatas – Ciência da Computação

Mineração, análise e classificação de dados é uma área emergente que tem ganhado bastante destaque no mundo empresarial e científico. Por isso, ferramentas devem ser testadas para adequada e eficiente preparação dos dados, acelerando a obtenção e discussão de resultados. Nessa pesquisa avaliou-se o uso da biblioteca Pandas para manipulação e análise de dados dos valores negociados na bolsa de valores de São Paulo no ano de 2017, com o objetivo de identificar por meios de gráficos da distribuição gaussiana os principais atributos para se utilizar em uma rede neural artificial como fator preditivo. A coleta de dados foi realizada no site BM&FBOVESPA. O arquivo, após tratamento dos dados, foi submetido na biblioteca Pandas e seu conteúdo transformado em uma tabela. Foi criada uma coluna extra para calcular o valor representado na curva gaussiana. A ferramenta foi adaptada para exibir graficamente a variação e comportamento dos dados de uma ou várias ações, simultaneamente; os mesmos dados também foram comparados após etapa de normalização demonstrando a eficiência da ferramenta testada. Assim, a biblioteca demonstrou ser capaz de realizar cálculos e demonstração de valores por meios de gráficos para os atributos escolhidos e facilitou a seleção dos mesmos para futura utilização em uma rede neural artificial.

Palavras-chave: Análise de dados; Pandas; Seleção de Atributos; Distribuição Gaussiana.