

Engenharia da Computação

Carlos Henrique Gomes dos Santos/1760092 /ENG 171i

Tema: Previsão de produtividade dos funcionários de vestuário

Birigui – SP

2021

Carlos Henrique Gomes dos Santos

Previsão de produtividade dos funcionários de vestuário

Relatório apresentado como requisito parcial para avaliação final – disciplina Mineração de Dados, do curso superior em Engenharia da Computação, ministrado pelo Prof Dr. Murilo Varges da Silva

Birigui – SP

2021

1. **Resumo**

Foi escolhido no site *UCI – Machine Learning Repository* a base de dados Previsão de produtividade dos funcionários de vestuário, onde a mesma deveria ter alguns dados ausentes, utilizando o algoritmo em linguagem *Python* fornecido pelo Prof. Dr Murilo para o processamento da base de dado escolhida, gerando dados estatísticos e filtrando dados da base que podem ser utilizadas para gerar informações relevantes para o processo.

Para isso então foi feito o download da base de dados e verificado quais atributos são relevantes, após isso é realizado o estudo do algoritmo fornecido, e a extração das informações com base nos dados fornecidos.

1. **Informações sobre a base de dados**

A Indústria de Vestuário é um dos principais exemplos da globalização industrial desta era moderna. É uma indústria altamente intensiva em mão-de-obra com muitos processos manuais. Satisfazer a enorme demanda global por produtos de vestuário depende principalmente do desempenho de produção e entrega dos funcionários nas empresas de fabricação de vestuário. Por isso, é altamente desejável entre os tomadores de decisão da indústria de vestuário acompanhar, analisar e prever o desempenho da produtividade das equipes de trabalho em suas fábricas.

1. **Informações de atributos**
2. date : Data em MM-DD-YYYY
3. day: Dia da Semana
4. quarter: Uma parte do mês. Um mês foi dividido em quatro partes
5. departament : Departamento associado com a intância
6. team\_no : Número de equipe associado com a intância
7. no\_of\_workers : Número de trabalhadores em cada equipe
8. no\_of\_style\_change : Número de mudanças no estilo de um determinado produto
9. targeted\_productivity : Produtividade direcionada definida pela Autoridade para cada equipe para cada dia.
10. smv : Valor do minuto padrão, é o tempo alocado para uma tarefa
11. wip : Trabalho em andamento. Inclui o número de itens inacabados para produtos
12. over\_time : Representa a quantidade de horas extras por cada equipe em minutos
13. incentive : Representa a quantidade de incentivo financeiro (no BDT) que permite ou motiva um determinado curso de ação.
14. idle\_time : A quantidade de tempo em que a produção foi interrompida por diversos motivos
15. idle\_men : O número de trabalhadores que estavam ociosos devido à interrupção da produção
16. actual\_productivity : O real % de produtividade que foi entregue pelos trabalhadores. Varia de 0-1.
17. Productivity: foi criado para se ter dados binários, actual\_productivity maior que 5, o campo recebe 1, se for menor recebe 0
18. **Funcionamento do algoritmo**

A primeira etapa do algoritmo é mostrar as 15 primeiras linhas da base de dados, logo após é mostrado informações gerais referentes a base nome de coluna, tipo de dado (*objetc, float, int*), após é mostrado a descrição de dados, quantidade de dados, qual valor vai substituir os que estão faltando (no nosso caso pela moda)*,* valor mínimo e máximo, desvio padrão, e alguns dados referente a média relacionadas a parte dos dados ¼ ½ e ¾ , é apresentado também a quantidade de dados faltantes em cada uma das colunas.

Podemos também filtrar dados de colunas especificas, gerando informações importantes, no nosso caso será utilizada a *survival.*

Quando se trata da substituição de dados ausentes para completar a base de dados, temos diversas opções, substitui por um número específico de escolha, pela mediana, média ou a moda

1. **Processamento da base de dados**