

UNIVERSIDADE FEDERAL DE SANTA MARIA CENTRO DE
TECNOLOGIA CURSO DE BACHARELADO EM CIÊNCIA DA
COMPUTAÇÃO

Arthur Bogacki Verissimo e Leandro Oliveira

TRABALHO 1: MINERAÇÃO DE DADOS DE UMA PADARIA

Santa Maria, RS
2024

RESUMO

TRABALHO 1: ATIVIDADES DE IMPLEMENTAÇÃO COM O MERGESORT

AUTORES: Arthur Bogacki Verissimo e Leandro Oliveira

PROFESSOR: Joaquim V C Assunção

Este relatório tem o objetivo de descrever os passos realizados na mineração de dados de uma padaria. O objetivo principal é realizar o devido pré-processamento de um arquivo **padaria_trab.json**, extrair as principais regras, encontrar o produto mais influente 1 para 1 e encontrar a regra dos produtos que implicam compra de “doce”

Relatório:

Pré-Processamento:

Formatando Arquivo: Como primeiro passo buscamos realizar o pré-processamento da tabela recebida. Como etapas do pré-processamento iniciamos abrindo o arquivo .json recebido e formatando o nome dos produtos de forma correta, convertemos o arquivo **.json** para um data frame utilizando a biblioteca **Pandas** do Python, removemos a coluna “compra” já que todos os valores são iguais “999” não seria útil para nossa mineração.

Explode: Usamos a função “explode” também da biblioteca Pandas para transformar uma coluna de um Data Frame que contém listas ou arrays em linhas, onde cada elemento da lista é expandido em uma linha

OneHot: Diversos algoritmos de machine-learning não podem trabalhar diretamente com dados categóricos, o One Hot transforma essas categorias em formato numérico onde cada categoria é apresentada por um vetor binário onde o bit ‘1’ é ativo e ‘0’ inativo

Exemplo:

| Pão | Queijo | Doce | Leite |
|-----|--------|------|-------|
| 1 | 0 | 0 | 1 |
| 0 | 0 | 1 | 1 |