



Universidade Federal
de São João del-Rei

TRABALHO PRÁTICO 2

Arthur Antunes Santos Silva

Este trabalho prático tem por objetivo exercitar conceitos e práticas iniciais da disciplina Mineração de dados. Foi proposto um tema que será estudado e aplicado técnicas de mineração de dados.

São João del Rei
Novembro de 2024

Sumário

1	Motivação	2
2	Problema a Ser Solucionado	2
3	Objetivo	2
4	Metodologia	3
4.1	Exploração Inicial e Pré processamento	3
4.2	Mineração de Dados	3
4.3	Pós-processamento e Avaliação dos resultados	3
5	Cronograma	4
6	Referências	4

1 Motivação

A produção agrícola é uma das principais atividades econômicas em diversos países, especialmente em regiões dependentes da agricultura para a geração de emprego e renda. A previsão da produção agrícola, considerando variáveis como clima, uso do solo, práticas agrícolas, entre outros, é crucial para a gestão eficiente de recursos, segurança alimentar e planejamento econômico. Além disso, as mudanças climáticas e a variabilidade nos padrões de produção agrícola aumentam a importância de modelar e prever essas variáveis com precisão.

A aplicação de técnicas de mineração de dados pode trazer grandes avanços na previsão de produtividade, permitindo que produtores e planejadores tomem decisões mais informadas e baseadas em dados. Com o avanço das tecnologias de aprendizado de máquina e análise de grandes volumes de dados, é possível identificar padrões complexos e preditivos que podem impactar diretamente a eficiência da produção agrícola e a sustentabilidade dos recursos.

2 Problema a Ser Solucionado

A variabilidade na produção agrícola, devido a fatores climáticos, econômicos e de manejo, dificulta a tomada de decisões eficientes e preditivas por agricultores e gestores. Os problemas centrais são:

1. Como prever com precisão os rendimentos agrícolas e identificar os fatores que mais impactam essa produtividade e como prever a produtividade agrícola com base em fatores históricos?
2. Como identificar variáveis críticas que impactam o rendimento?

Isso inclui explorar relações entre variáveis climáticas, sazonais, de manejo e rendimentos, e criar modelos preditivos úteis para decisões estratégicas.

3 Objetivo

O objetivo principal do projeto é aplicar técnicas de mineração de dados para realizar análises preditivas e identificar padrões relevantes em um dataset de produção agrícola disponível. Especificamente:

1. Análise Exploratória: Identificar padrões e tendências nos dados de produção agrícola, como sazonalidade ou correlação com variáveis climáticas.
2. Modelagem Preditiva: Desenvolver modelos para prever a produtividade agrícola com base em variáveis históricas.
3. Identificação de Fatores Críticos: Determinar as variáveis que mais impactam a produtividade, auxiliando na priorização de esforços e recursos.

4 Metodologia

A metodologia poderá ser dividida nas seguintes etapas:

4.1 Exploração Inicial e Pré processamento

1. Coleta e Compreensão dos Dados: Análise exploratória do dataset para entender sua estrutura, variáveis disponíveis e identificar possíveis inconsistências ou valores ausentes.
2. Limpeza: Tratamento de valores ausentes, remoção de duplicatas e correção de inconsistências.
3. Transformação: Normalização de variáveis e codificação de variáveis categóricas.

4.2 Mineração de Dados

1. Modelagem Preditiva: Aplicação de algoritmos como regressão linear, árvores de decisão e redes neurais para prever rendimentos agrícolas.
2. Agrupamento (Clustering): Utilização de técnicas como K-means e DBSCAN para segmentar dados e identificar grupos com características semelhantes.

4.3 Pós-processamento e Avaliação dos resultados

1. Validação dos Modelos: Utilização de métricas como desvio padrão, MAE e coeficiente de determinação para avaliar a performance dos modelos preditivos.
2. Visualização: Apresentar os resultados de forma clara, com gráficos e permitindo a interpretação prática.

5 Cronograma

Entrega	Atividade	Prazo
Proposta	Levantamento do tema do trabalho	21/11/2025
Exploração Inicial e Pré processamento	Limpeza, normalização, transformação e análise inicial do dataset, visualização e análise estatística descritiva dos dados.	01/12/2024 - 31/12/2024
Apresentações parciais	Criação e entrega da apresentação parcial com os primeiros resultados.	14/01/2025
Desenvolvimento do Modelo Preditivo	Construção dos modelos preditivos com técnicas de mineração de dados.	31/01/2025
Avaliação e Validação do Modelo	Criação de visualizações e relatórios para interpretar os resultados.	01/02/2025
Visualização e Interpretação	Criação de visualizações e relatórios para interpretar os resultados.	02/02/2025
Apresentação Final	Apresentação detalhada de resultados, metodologia e conclusões.	06/02/2025 - 11/02/2025
Escrita do Artigo Científico	Redação do artigo em LaTeX, incluindo Introdução, Trabalhos Relacionados, Metodologia, Resultados e Conclusões.	11/02/2025

Figura 1: Cronograma proposto para realização das atividades

6 Referências

- [1] <https://www.kaggle.com/datasets/imtkaggleteam/crop-production>
- [2] Ortiz-Bobea, Ariel, et al. "The historical impact of anthropogenic climate change on global agricultural productivity." arXiv preprint arXiv:2007.10415 (2020). Acesso em: 18 nov. 2024.
- [3] SILVA, José Graziano da; ALVES, Eliseu; OSAKI, Mauro. Produção agrícola mundial: o potencial do Brasil. Brasília: EMBRAPA, 2016. Disponível em: <https://ainfo.cnptia.embrapa.br/arquivos/pdf/producao_agricola-mundial.pdf>. Acesso em: 19 nov. 2024.