

PROTÓTIPO DE UM SISTEMA DE RECOMENDAÇÃO DE DADOS AGRÍCOLAS BASEADO EM UMA PLATAFORMA DE DADOS ESCALÁVEL

1 INTRODUÇÃO

Transformação digital consolidou os dados como recurso estratégico do século XXI, essenciais para compreender comportamentos, identificar tendências e apoiar a tomada de decisão em diversos setores. O conceito de Big Data, entendido como o conjunto de métodos e tecnologias voltados à coleta, armazenamento, processamento e análise de grandes volumes de informações, tem sido amplamente explorado por empresas de tecnologia, como Google, Facebook e Amazon, que processam quantidades massivas de dados para responder rapidamente às mudanças de mercado e gerar novos conhecimentos (KLEPPMANN, 2017).

No contexto agrícola, a fragmentação das informações em múltiplas bases dificulta análises integradas e compromete a construção de modelos confiáveis de previsão e recomendação. Pesquisas apontam que a integração de dados meteorológicos, de manejo e de cultivares pode elevar significativamente a produtividade e reduzir riscos associados à instabilidade climática (KAMILARIS; KARTAKOULLIS; PRENAFETA-BOLDÚ, 2017; BRONSON; KNEZEVIC, 2016). Dessa forma, torna-se fundamental o desenvolvimento de plataformas que consolidem essas informações, garantindo escalabilidade, acessibilidade e suporte a decisões mais embasadas no setor.

Este trabalho tem como objetivo propor um protótipo de sistema de recomendação de dados agrícolas baseado em arquitetura de Big Data e serviços em nuvem. A proposta centraliza informações meteorológicas e agronômicas, processa-as em pipelines distribuídos e disponibiliza insights por meio de dashboards interativos, com foco na análise da soma térmica e da pluviometria na produtividade do milho. Dessa maneira, a plataforma contribui para decisões mais assertivas de agrônomos, pesquisadores e produtores, promovendo maior eficiência no uso de recursos e sustentabilidade na produção agrícola.

2 FUNDAMENTAÇÃO TEÓRICA

2.1 AGRICULTURA SUSTENTÁVEL E DE PRECISÃO

O aumento da demanda global por alimentos intensificou a busca por práticas que conciliem produtividade e preservação ambiental. A Agricultura Sustentável, consolidada a partir dos anos 1980, propõe sistemas produtivos que mantenham a viabilidade econômica e reduzam impactos ecológicos (EDWARDS, 2020). Nesse cenário, a Agricultura de Precisão utiliza sensores, GPS e Internet das Coisas (IoT) para monitorar a variabilidade do solo e das plantações, otimizando o uso de insumos e promovendo maior eficiência (AKHTER; SOFI, 2022).

2.2 BIG DATA E AGRICULTURA

O conceito de Big Data Analytics refere-se à capacidade de processar grandes volumes de dados estruturados e não estruturados, extraíndo padrões e gerando insights estratégicos (WAGA; RABAH, 2014; BRONSON; KNEZEVIC, 2016). Na agricultura, seu uso permite integrar informações meteorológicas, de solo e de cultivares, reduzindo riscos e orientando decisões mais assertivas (KAMILARIS; KARTAKOULLIS; PRENAFETA-BOLDÚ, 2017).

2.3 COMPUTAÇÃO EM NUVEM

A computação em nuvem possibilita a escalabilidade necessária para o armazenamento e processamento de dados agrícolas em larga escala, com alta disponibilidade e menor custo de infraestrutura (BHATTARAI et al., 2019; AMAZON, 2023). Esse modelo garante resiliência e segurança, além de permitir a integração de múltiplas fontes de dados em ambientes distribuídos.

2.4 TECNOLOGIAS DE PROCESSAMENTO E VISUALIZAÇÃO

O Apache Spark destaca-se pelo processamento paralelo e distribuído de grandes volumes de dados, reduzindo o tempo de resposta e ampliando a eficiência analítica (ZAHARIA et al., 2016). Em complemento, o Apache Airflow oferece orquestração de pipelines de coleta, transformação e carga, assegurando automação e escalabilidade (HARENSLAK; RUITER, 2021). Por fim, o Microsoft Power BI atua como ferramenta de análise e visualização, transformando dados processados em dashboards interativos que apoiam a tomada de decisão (MICROSOFT, 2023).

3 METODOLOGIA

O desenvolvimento do protótipo seguiu uma abordagem em nuvem, estruturada em etapas que compreendem a coleta, o armazenamento, o processamento, a catalogação e a visualização dos dados.

Na primeira etapa, realizou-se o mapeamento e a coleta das fontes de dados, que incluíram registros meteorológicos do Instituto Nacional de Meteorologia (INMET), dados climatológicos da API NASA POWER e informações agronômicas do Zoneamento Agrícola de Risco Climático (ZARC/MAPA). Essas bases foram integradas em um data lake construído no Amazon S3, organizado em camadas raw, curated e analytics, permitindo escalabilidade e segurança no armazenamento.

Na segunda etapa, os dados foram submetidos a processamento distribuído com o Apache Spark, responsável por aplicar transformações e preparar os conjuntos para análise. O Apache Airflow foi utilizado para orquestração dos pipelines, automatizando a execução de cargas em lote e o monitoramento dos fluxos.

Em seguida, a etapa de catalogação e governança foi implementada com o AWS Glue, assegurando padronização de esquemas e integração com o AWS Athena para consultas em linguagem SQL. Esse processo viabilizou análises otimizadas, permitindo maior confiabilidade e rastreabilidade dos dados.

Por fim, os resultados processados foram disponibilizados em dashboards interativos no Microsoft Power BI, que possibilitaram a visualização de indicadores como soma térmica, precipitação acumulada e riscos climáticos. Essa interface intuitiva facilita a análise de cenários e apoia a tomada de decisão técnica no setor agrícola.

A Figura 1 apresenta uma visão simplificada da arquitetura metodológica adotada, que integra os serviços de nuvem à camada de análise de dados.

Figura 1 - Fluxo metodológico do protótipo de sistema de recomendação



Fonte: Dados da pesquisa, 2023.

A ilustração sintetiza o caminho percorrido pelos dados, desde sua coleta em fontes meteorológicas e agrônômicas até a disponibilização em dashboards analíticos. Observa-se que a metodologia adotada favorece escalabilidade, automação e integração de diferentes serviços, compondo uma arquitetura resiliente e voltada às necessidades da agricultura de precisão.

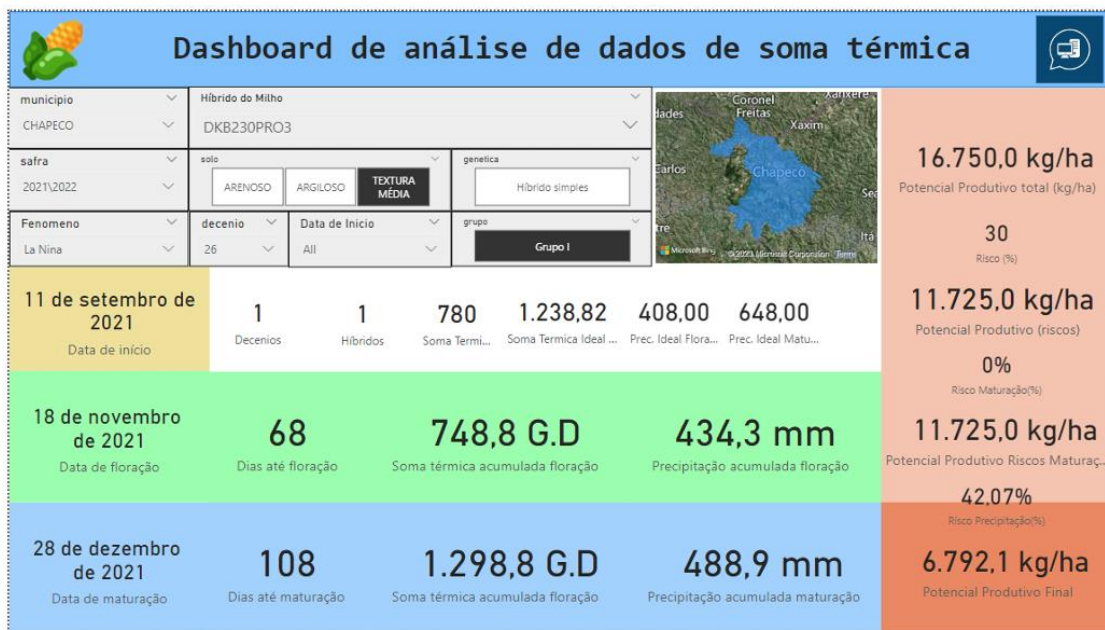
4 ANÁLISE E DISCUSSÃO DOS RESULTADOS

A análise concentrou-se inicialmente na soma térmica do milho, indicador essencial para estimar o ciclo de desenvolvimento da planta e sua produtividade. De acordo com Wagner et al. (2013), o intervalo ideal para o crescimento do milho está entre 24 °C e 30 °C, sendo que temperaturas extremas comprometem o acúmulo energético necessário para as fases fenológicas. O sistema proposto permitiu calcular a soma térmica acumulada em diferentes localidades, cruzando dados meteorológicos do INMET e do ZARC, com recortes temporais ajustados por safra e decênio.

Em complemento, foram integradas informações meteorológicas do INMET (coletas horárias e históricas desde 2000), dados da API NASA POWER e informações de risco climático disponibilizadas pelo MAPA. Essa combinação ampliou a robustez das análises, permitindo projeções mais confiáveis sobre produtividade e riscos associados a fenômenos climáticos como El Niño e La Niña.

A Figura 2 apresenta a visão granular do dashboard, que permite filtrar híbridos específicos, localidades e períodos, retornando dados de soma térmica e pluviometria. Essa visualização detalhada possibilita análises precisas em cenários definidos pelo usuário.

Figura 2 - Dashboard: visão granular de híbridos e condições climáticas

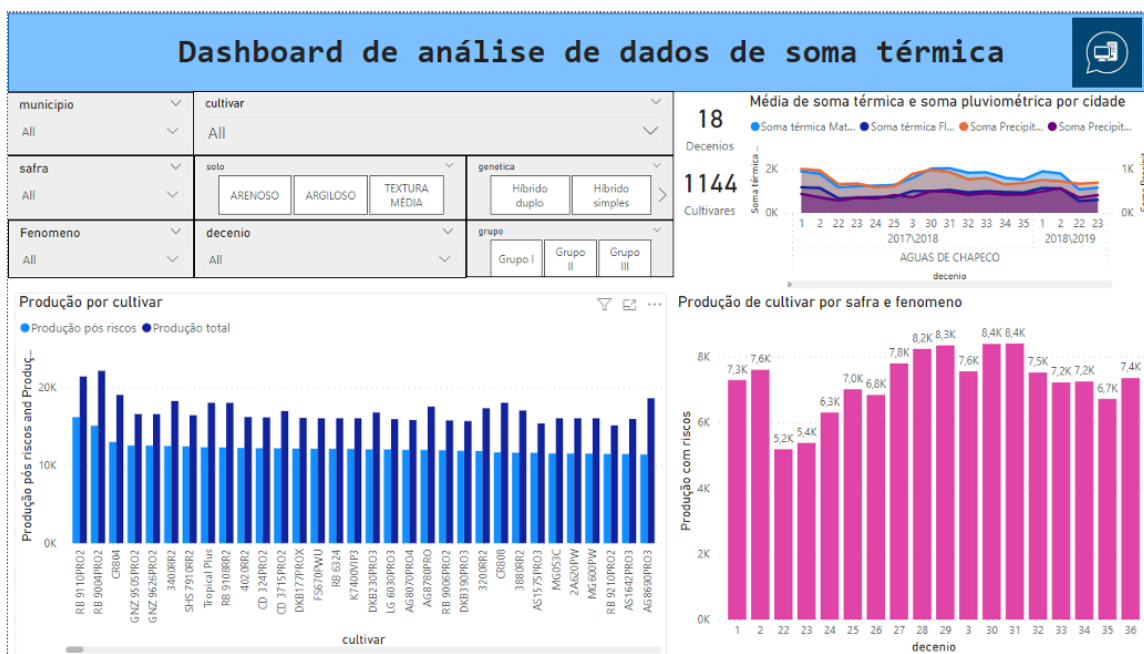


Fonte: Dados da pesquisa, 2023.

A visualização granular demonstrou resultados consistentes com avaliações de campo, apresentando diferença média de apenas 1% no risco climático e 5% no potencial produtivo quando comparada a medições realizadas por técnicos e entidades como a EPAGRI. Esses resultados validam a confiabilidade do sistema como ferramenta de apoio à decisão agrícola.

A Figura 3 mostra a visão agrupada por híbridos, que compara o desempenho de diferentes cultivares em localidades e safras específicas, permitindo identificar aqueles com melhor adaptação às condições climáticas.

Figura 3 - Dashboard: visão agrupada de cultivares e riscos climáticos



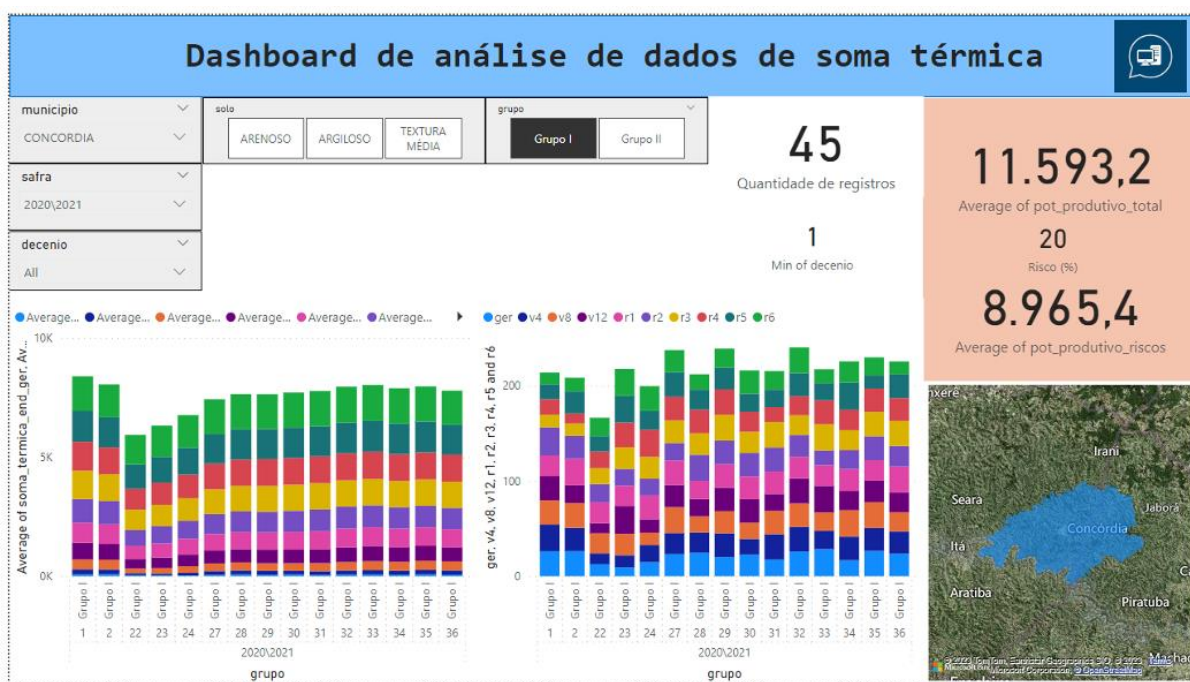
Fonte: Dados da pesquisa, 2023.

Essa análise agrupada demonstrou potencial de uso para pesquisas e recomendações agronômicas, pois facilita a comparação entre diferentes cultivares sob condições ambientais variadas. Além disso, a visualização por fases fenológicas destacou o impacto da precipitação e da soma térmica em cada etapa do ciclo da planta, oferecendo subsídios técnicos para o manejo de híbridos em períodos críticos.

Portanto, os resultados evidenciam que a plataforma desenvolvida é capaz de integrar múltiplas fontes de dados e disponibilizá-los em formato analítico, validando sua aplicabilidade tanto para fins científicos quanto para o apoio a decisões práticas de agrônomos e produtores.

Além das análises granulares e agrupadas, a plataforma disponibilizou uma visualização focada nas etapas fenológicas do milho, permitindo acompanhar o acúmulo de soma térmica e a precipitação em cada fase do ciclo da planta. Essa visão, apresentada na Figura 4, possibilita identificar períodos críticos de desenvolvimento e avaliar o impacto de condições climáticas específicas sobre o desempenho do cultivo.

Figura 4 - Dashboard: Visão Agrupada das etapas fenológicas do milho



Fonte: Dados da pesquisa, 2023.

5 CONSIDERAÇÕES FINAIS

A construção e implementação do protótipo evidenciaram sua relevância como ferramenta de apoio à tomada de decisão no setor agrícola. A plataforma demonstrou capacidade de integrar múltiplas fontes de dados, como registros meteorológicos e informações de cultivares, disponibilizando análises de forma escalável, confiável e acessível. Os resultados validaram a aplicabilidade da solução, com alto grau de proximidade em relação a dados de campo, reforçando sua utilidade prática para agrônomos e pesquisadores.

Apesar dos avanços, algumas limitações foram identificadas, como a impossibilidade de integrar em tempo real determinados bancos relacionais devido a restrições de segurança e

custo, além do desafio no acesso a bases públicas com uso restrito ou caráter confidencial. Ainda assim, os objetivos centrais foram alcançados, demonstrando a viabilidade técnica do sistema.

Conclui-se que a plataforma desenvolvida representa um passo significativo para o uso estratégico de Big Data na agricultura, oferecendo base sólida para futuras aplicações, como o uso de dados em tempo real e a incorporação de algoritmos de aprendizado de máquina. Dessa forma, o trabalho contribui para a modernização das práticas agrícolas, reforçando a importância da tecnologia como aliada da sustentabilidade e da produtividade no campo.

REFERÊNCIAS

AKHTER, R.; SOFI, S. A. Precision agriculture using IoT data analytics and machine learning. **Journal of King Saud University – Computer and Information Sciences**, v. 34, n. 8, p. 5602–5618, 2022.

AMAZON. AWS Documentation. 2023. Disponível em: <https://docs.aws.amazon.com/>. Acesso em: 15 ago. 2025.

BHATTARAI, B. P. et al. Big data analytics in smart grids: state-of-the-art, challenges, opportunities, and future directions. **IET Smart Grid**, v. 2, p. 141–154, 2019.

BRONSON, K.; KNEZEVIC, I. Big data in food and agriculture. **Big Data & Society**, v. 3, n. 1, p. 1–12, 2016.

EDWARDS, C. A. **Sustainable agricultural systems**. Boca Raton: CRC Press, 2020.

HARENSLAK, Bas P.; DE RUITER, Julian. **Pipelines de dados com Apache Airflow**. Simon e Schuster, 2021.

KAMILARIS, A.; KARTAKOULLIS, A.; PRENAFETA-BOLDÚ, F. X. A review on the practice of big data analysis in agriculture. **Computers and Electronics in Agriculture**, v. 143, p. 23–37, 2017.

KLEPPMANN, M. *Designing data-intensive applications*. Beijing: O’Reilly, 2017.

MICROSOFT. Microsoft Documentation. 2023. Disponível em: <https://learn.microsoft.com/>. Acesso em: 15 ago. 2025.

WAGA, D.; RABAH, K. Environmental conditions’ big data management and cloud computing analytics for sustainable agriculture. **World Journal of Computer Application and Technology**, v. 2, p. 73–81, 2014.

WAGNER, M. V. et al. Estimativa da produtividade do milho em função da disponibilidade hídrica em Guarapuava, PR, Brasil. **Revista Brasileira de Engenharia Agrícola e Ambiental**, v. 17, n. 2, p. 170–179, 2013.

ZAHARIA, M. et al. Apache Spark: a unified engine for big data processing. **Communications of the ACM**, v. 59, n. 11, p. 56–65, 2016.