Introdução à Ciência de Dados

usando Árvores

Enzo Benedetto Proença, 10418579

Gabriel Alves de Freitas Spinola Sucupira, 10418133

Lucas Pires de Camargo Sarai, 10418013

Vitor Alves Pereira, 10410862

Sumário

[Introdução 3](#_Toc183023227)

[Descrição do Dataset 4](#_Toc183023228)

[Modelagem nas Árvores 5](#_Toc183023229)

[Descrição da Implementação das Estruturas de Dados e as Operações Realizadas 6](#_Toc183023230)

[Perguntas Exploratórias 7](#_Toc183023231)

[Comparação de Desempenho das Árvores BST e AVL 12](#_Toc183023232)

[Conclusão 13](#_Toc183023233)

[Referências 14](#_Toc183023234)

Introdução

**O que são?**

De acordo com a definição no site do Programa das Nações Unidas para o Desenvolvimento: “Os Objetivos de Desenvolvimento Sustentável (ODS), também conhecidos como Objetivos Globais, foram adotados pelos países-membros das Nações Unidas em 2015 como um apelo universal à ação para acabar com a pobreza, proteger o planeta e garantir que até 2030 todas as pessoas desfrutem de paz e prosperidade”.

Em suma, são objetivos variados definidos em reuniões entre os países membros para buscarem desenvolver e preservar o planeta.

**Objetivo escolhido**

Um desses objetivos é o 8: “Trabalho decente e crescimento econômico”. E nosso trabalho será focado nesse tópico, onde trabalhamos um banco de dados sobre a variação do PIB em todos os setores entre 2002 e 2024.

Descrição do Dataset

O *Dataset* escolhido foi “Taxas de crescimento do Valor Adicionado, por setores e subsetores de atividade econômica, dos Impostos Líquidos de Subsídios e do Produto Interno Bruto, no mês em relação ao mesmo mês do ano anterior”. Em outras palavras, mostra a variação percentual do PIB por setor econômico de um mês em relação ao mesmo mês do ano anterior. Ou seja, se janeiro de 2003 possui uma variação de +15,8% no setor da agropecuária, isso indica que em relação a janeiro de 2002, esse setor do PIB aumentou em 16%, aproximadamente.

Dentre os setores, optamos por analisar somente a agropecuária, visto que poderíamos nos aprofundar mais se focássemos em apenas um, e escolhemos esse setor mais especificamente por variar bastante e possuir vários fatos que podem ser analisados.

Abaixo, segue trecho da base de dados:

Interface gráfica do usuário, Tabela

Descrição gerada automaticamente

Modelagem nas Árvores

Para cada árvore, desenvolvemos dois tipos de organização dos dados: pela taxa de crescimento do setor agropecuário e pelas datas mês/ano de cada dado.

Texto

Descrição gerada automaticamente

Para o primeiro, usamos a taxa de crescimento por não ter valores repetidos, visto que são valores do tipo *double* e possuem muitas casas decimais. Seguindo a lógica de BST que é aplicada também na AVL, as taxas de crescimento menores que uma taxa em específico estarão à sua esquerda, enquanto as outras estarão à direita. Esse tipo de árvore nos auxiliou a descobrir os *outliers*, ou seja, as taxas de crescimento mais distantes do padrão esperado, visto que ao percorrer até o filho mais distante à esquerda, encontramos o menor valor, e repetindo o processo na direção oposta, encontramos o maior valor.

Para o segundo tipo, usamos o ano e o mês, nessa ordem, para organizar os valores. Então as operações começavam procurando pelo ano, sabendo que anos menores estariam à esquerda, enquanto anos maiores, à direita. No entanto, temos vários valores com o mesmo ano, só que em meses diferentes; Então, usando a ideia de “chave primária composta” de banco de dados, onde um valor é diferenciado de outro pela combinação de chaves primárias e não por apenas uma única, usamos o mês para diferenciar valores de mesmo ano, utilizando o valor inteiro representativo de cada mês (01 para janeiro, 02 para fevereiro etc.), visto que se usássemos o valor texto como “Janeiro” ou “Fevereiro”, os dados seriam ordenados pela ordem lexicográfica, logo “Abril”, por exemplo, seria tratado como o primeiro mês do ano, por começar com “A”. Assim sendo, os valores à esquerda de um nó possuem não somente o valor de ano menor ou igual ao desse nó, como possuem o valor de mês menor. Esse segundo tipo nos auxiliou a identificar as médias anuais de anos específicos, pois tínhamos como procurar de forma eficiente os dados de acordo com a data.

Descrição da Implementação das Estruturas de Dados e as Operações Realizadas

Perguntas Exploratórias

Texto

Descrição gerada automaticamente

• **Pergunta Exploratória 1**: *“Qual foi o motivo do maior crescimento percentual em todo o período (2003-2024)?*

Em todo o período de 2003 a 2024, o maior crescimento no estado de São Paulo ocorreu entre 2014 e 2016, período em que houve uma recessão econômica em nível nacional, porém o setor agropecuário paulista conseguiu crescer por diversos fatores: diversificação na produção, liderança na produção de cana-de-açúcar e etanol no país e desvalorização do real são alguns fatores.

O primeiro fator auxiliou o estado a compensar a queda em alguns produtos agrícolas com o crescimento na produção de outros, se mantendo competitivo em nível nacional e internacional.

A liderança na produção de cana e etanol também foi importante, pois esses produtos se mantiveram em alta no período, e sendo o estado o líder na produção, por consequência, continuou crescendo.

Com a desvalorização do real frente ao dólar, a venda de produtos em dólar aumentou os ganhos em real, além de os produtos se tornarem mais baratos e, assim, mais procurados e competitivos em nível internacional.

Portanto, mesmo com o PIB geral em queda, o setor cresceu muito no estado durante esse período, como podemos ver nos gráficos abaixo:

Gráfico, Gráfico de linhas

Descrição gerada automaticamente

Gráfico

Descrição gerada automaticamente com confiança média

• **Pergunta Exploratória 2**: *“Qual foi o motivo da maior queda em todo o período (2003-2024)?*

A maior queda ocorreu entre 2008 e 2009, com consequências até 2010, ano que teve uma taxa de fruto da crise financeira global de 2008-2009 que começou nos Estados Unidos e ficou conhecida como “Grande Recessão”. Essa crise começou no setor imobiliário norte-americano ao qual nos anos 2000 teve uma expansão muito grande no mercado, com preços de imóveis em rápida crescente e oferta excessiva de hipotecas de alto risco. Eventualmente, essa bolha imobiliária explodiu, levando à falência do banco Lehman Brothers, um banco tão importante que sua quebra levou os efeitos para fora dos EUA e a crise pôde ser sentida em todo mundo.

As consequências no setor agropecuário mundial foram devastadoras, principalmente pela queda da demanda global de *commodities* (matérias-primas) agrícolas, além da inflação de seus preços, somados a desconfiança com a contratação de crédito agrícola, visto que a crise começou com a contração excessiva de crédito, gerando uma desconfiança global nesse meio, dificultando o financiamento para insumos, equipamentos e expansão da produção.

Esse cenário afetou muito o estado de São Paulo, como podemos ver nos gráficos abaixo:

Gráfico, Gráfico de linhas

Descrição gerada automaticamente

Gráfico, Gráfico de cascata

Descrição gerada automaticamente

Por outro lado, importante notar que a média geral de crescimento do PIB em São Paulo foi a mais elevada em 2010, dado que a partir desse ano a economia nos outros setores começou a se recuperar e como os valores nos outros anos estavam numa queda muito grande, um mínimo crescimento que fosse apresentado acabaria subindo muito a curva do gráfico.

• **Pergunta Exploratória 3**: *“Existiram valores fora do esperado (outliers)? Qual o motivo?”*

*Gráfico, Gráfico de cascata

Descrição gerada automaticamente*

*Gráfico

Descrição gerada automaticamente*

Em 2017, houve valores muito discrepantes em maio e junho em relação ao crescimento que o setor vinha tendo: -25% e -34,41%, respectivamente¨. Isso ocorreu por conta de condições climáticas desfavoráveis, principalmente com a escassez de chuvas, e essa seca afetou a produção agrícola, incluindo uma das principais culturas agrícolas de São Paulo, que vinha sendo uma das responsáveis pelo crescimento do setor agrícola mesmo durante a crise: a cana-de-açúcar. Com a queda nessa produção, esse setor ficou mais vulnerável às oscilações econômicas que o PIB do estado vinha sofrendo no ano de 2017.

Por outro lado, em 2015 teve uma alta considerável em março e junho, atingindo valores de +30% e +34% em relação ao ano anterior, respectivamente. Nesses meses, o crescimento é acentuado pela colheita de grãos, onde se intensifica essa colheita antes de começar o inverno, principalmente das culturas de soja e milho, sendo março o início e junho o fim, por isso uma variação maior nesses meses, além de junho costuma ser marcado pela época de abate de gado e variação nos preços da carne.

Logo, nota-se que esses dois meses costumam atrair valores *outliers.* Isso porque eles são os meses de rotação de culturas animais e vegetais e, portanto, geram uma oscilação no PIB agropecuário.

• **Pergunta Exploratória 4**: *“Qual foi o crescimento/decrescimento da participação do setor agropecuário no período pré-pandemia?”*

*Gráfico, Gráfico de linhas

Descrição gerada automaticamente*

Como verifica-se ao analisar o gráfico e pelos cálculos de variação percentual entre 2017 e 2018 e entre 2018 e 2019, nota-se que houve uma queda entre 2017 e 2018 de -9,63%, porém cresceu 6.83% entre 2018 e 2019, com uma média no período desses três anos de -1.40%. A queda inicial pode ser explicada pelo fato de ter existido um crescimento muito grande nos anos anteriores, e uma produção um pouco mais regular, e com quedas bruscas nos meses de março e junho (ver questão 3), ocasionaram uma queda anual percentual elevada em relação ao ano anterior, além de fatores como condições climáticas adversas (secas, como a do alto do Rio Negro) e a greve dos caminhoneiros do final de 2017 pelo preço dos combustíveis, custos operacionais e pela insatisfação com as condições de trabalho, consequentemente paralisando o transporte de insumos agrícolas para os portos e para as cadeias de produção.

Já no período de 2018 e 2019 houve a eleição do novo presidente, que investiu no setor agropecuário por meio do Plano Safra, destinando recursos para o custeio e comercialização de produtos agrícolas, e para investimentos, como a compra de maquinário.

• **Pergunta Exploratória 5**: *“Durante a pandemia, a economia global sofreu um colapso. Como ela afetou o setor agropecuário paulista?*

*Gráfico, Gráfico de linhas

Descrição gerada automaticamente*

Com a crise global encadeada pela pandemia, o setor agropecuário foi muito afetado principalmente entre 2020 e 2021, no início da pandemia, com uma queda de -7,81%. Entretando, a economia do setor se recuperou no final da pandemia entre 2021 e 2022 com um crescimento de 5,24%. Os motivos da queda foram maiores para os pequenos produtores pelas dificuldades de manutenção da dinâmica produtiva e comercial; como viajar para buscar e vender os produtos e impactos nos volumes de produção pela redução da equipe em serviço em nome do isolamento. Assim, a variação média percentual foi de -1,28%, representando uma queda menor que entre 2017 e 2019, o que pode ser explicado, principalmente, pelo investimento do novo governo.

Comparação de Desempenho das Árvores BST e AVL

Comparação do desempenho entre a BST e AVL, com gráficos e discussão sobre as diferenças observadas. Discutir sobre a eficácia de cada implementação de árvore para o tipo de análise realizada.

Conclusão

Conclusão do trabalho com uma discussão sobre as implicações dos resultados e as limitações do estudo.

Referências

<https://economia.uol.com.br/noticias/redacao/2024/07/04/o-ano-em-que-o-mundo-quebrou-entenda-a-crise-financeira-de-2008.htm>

<https://brasilescola.uol.com.br/geografia/crise-financeira-global.htm>

<https://agenciabrasil.ebc.com.br/economia/noticia/2018-12/seca-atingiu-38-milhoes-de-brasileiros-em-2017-e-cheias-2-milhoes#:~:text=Seca%20atingiu%2038%20milh%C3%B5es%20de%20brasileiros%20em%202017%20e%20cheias%2C%202%20milh%C3%B5es,-Estudo%20da%20ANA&text=No%20ano%20passado%2C%20quase%2038,afetados%20por%20cheias%20e%20inunda%C3%A7%C3%B5es>.

<https://veja.abril.com.br/economia/falta-de-chuva-pode-elevar-em-mais-de-10-a-conta-de-luz>

<https://ojs.brazilianjournals.com.br/ojs/index.php/BRJD/article/view/2759>

<https://www.google.com/url?sa=i&url=https%3A%2F%2Fwww.gov.br%2Fagricultura%2Fpt-br%2Fassuntos%2Fnoticias-2022%2Fplano-safra-disponibiliza-r-340-8-bilhoes-para-o-setor-agropecuario&psig=AOvVaw03JVhEJpBg7BMTsN4lnNu3&ust=1732392410824000&source=images&cd=vfe&opi=89978449&ved=0CAYQrpoMahcKEwjAmPLd3vCJAxUAAAAAHQAAAAAQBA>

<https://agenciabrasil.ebc.com.br/economia/noticia/2022-06/plano-safra-20222023-anuncia-r-3408-bilhoes-para-agropecuaria#:~:text=Sustentabilidade,monitoramento%20da%20umidade%20no%20solo>

<https://www.cartacapital.com.br/economia/plano-safra-governo-preve-r-400-bilhoes-em-creditos-para-medios-e-grandes-produtores-rurais/#:~:text=Do%20total%2C%20quase%20240%20bilh%C3%B5es,como%20a%20compra%20de%20maquin%C3%A1rio>.

<https://www1.folha.uol.com.br/mercado/2018/09/entenda-a-crise-dos-caminhoneiros.shtml>

<https://www.infoescola.com/historia/greve-dos-caminhoneiros-2018/>

<https://brasil.elpais.com/brasil/2018/05/24/economia/1527177800_693499.html>

<https://economia.uol.com.br/noticias/bbc/2018/05/30/greve-dos-caminhoneiros-a-cronologia-dos-10-dias-que-pararam-o-brasil.htm>