

Aplicação de Aprendizado de Máquina para Predição do Preço da Cesta Básica

Romário Getúlio Vargas
Curso de Ciência da Computação
Universidade do Vale do Itajaí -
UNIVALI
romario.getulio@gmail.com

Fernanda dos Santos Cunha
Curso de Ciência da Computação
Universidade do Vale do Itajaí -
UNIVALI
fernanda.cunha@univali.br

Anita Maria da Rocha Fernandes
Laboratório de Inteligência Aplicada - LIA
Universidade do Vale do Itajaí - UNIVALI
anita.fernandes@univali.br

ABSTRACT

Four foods from the Brazilian Basic Basket were chosen for KDD application to predict future prices with their trends and causes, if they are related to the data sources used in the research. To predict prices, the Holt-Winters method was used, and to define the trend, K-Means was used. The tools used in this development were Jupyter Notebook, Anaconda, Excel, PostgreSQL, Docker, ZUDA Orange3 and the Python language, with its main libraries related to data science, such as Pandas, NumPy, Matplotlib and Seaborn.

KEYWORDS

Aprendizado de Máquina, Modelos Preditivos, Cesta Básica.

1 Introdução

A baixa renda atinge muitas famílias brasileiras que não têm acesso para obter o mínimo de necessidades básicas para viver uma vida digna. Para melhorar as condições das famílias brasileiras o governo tem adotado programas de transferência de renda com condicionalidades, tendo como objetivo central reduzir a pobreza e as desigualdades sociais. O Programa Bolsa Família (PBF) e o que está ativo, ele foi criado em 2004 como resultado da unificação de programas sociais já existentes – Bolsa Escola, Vale Alimentação, Bolsa Alimentação e Auxílio Gás [1]. Segundo um estudo realizado pela Empresa Brasileira de Pesquisa Agropecuária Ministério da Agricultura, Pecuária e Abastecimento – EMBRAPA [2], ao avaliar os possíveis cenários da mudança do clima e os impactos relacionados à agricultura familiar nas regiões Norte e Nordeste do Brasil, percebeu-se, fundamentalmente, que os pequenos e médios agricultores terão de se adaptar a um mundo de crescente variabilidade climática. Como consequência, estas variações podem gerar perdas nas produções dos alimentos, fazendo com que os preços se elevem.

O estudo ainda aponta que as emissões antropogênicas de Gases de Efeito Estufa (GEE) têm aumentado desde a era pré-industrial, principalmente impulsionadas pelo crescimento econômico e populacional, e atualmente são maiores que nunca. De 1750 a 2013, a concentração de CO₂ aumentou de 280ppm (partes por milhão) para 400ppm. Por conseguinte,

as últimas três décadas têm sido mais quentes que qualquer década anterior desde 1850, enquanto as temperaturas das superfícies terrestres e oceânicas combinadas aumentaram, aproximadamente, 0,85°C de 1880 a 2012 [2]. Conforme apresentado, são diversos os fatores que podem influenciar nas precificações dos produtos da Cesta Básica, e um deles é a elevação da temperatura terrestre, onde o aumento de dias de calor pode ocasionar perdas muito grandes na produção desses alimentos. Devido a estes fatores, é cada vez mais importante a realização de pesquisas nesta área.

Neste sentido, este artigo apresenta a aplicação dos métodos Holt-Winters e K-Means para predição do preço da cesta básica no Brasil. A pesquisa buscou identificar valores futuros dos preços de quatro alimentos da Cesta Básica: arroz, feijão, batata inglesa e tomate. As tecnologias utilizadas foram ZUDA Orange3, PostgreSQL, Excel, DBeaver, Jupyter Notebook, Anaconda e a linguagem Python, com suas bibliotecas Pandas, NumPy, Matplotlib e Seaborn. Tal escolha baseou-se no fato de a Jupyter Notebook com a linguagem Python fornecerem um ambiente propício para computação estatística e gráfica.

2 Métodos Utilizados

Foram aplicados dois modelos de predição: Holt-Winters e K-Means. A seguir tem-se uma breve descrição de ambos.

O método Holt-Winters e suas variações são usados de forma adequada para previsão de dados de entrada que mostram tendências e comportamentos sazonais. O algoritmo é uma das técnicas de previsão mais populares para séries temporais. Apesar de existir há décadas, ele ainda é muito utilizado para fins de detecção de anomalias e, especialmente, na previsão de tempo. Esse modelo é uma extensão do modelo de Holt (suavização exponencial dupla), desenvolvido por Winter. O método pode lidar com muitos padrões sazonais complicados, simplesmente encontrando o valor central e adicionando os efeitos de inclinação e sazonalidade [3].

K-Means é um método de segregar em torno de centros (centroides) diversos dados, ou seja, realizar a clusterização, gerando o efeito de particionar n observações dentre k grupos onde cada observação pertence ao grupo mais próximo da média, resultando em uma divisão do espaço de dados [4].

3 Metodologia

Para iniciar a pesquisa, foi necessário criar uma base única considerando dados oriundos do IBGE, EPAGRI/CIRAN, e DIEESE. O elemento agregador destas bases foi o código do IBGE dos municípios. Utilizou-se o Excel para um tratamento manual dos dados, e criou-se rotinas em Python no Jupyter Notebook para normalizá-los, padronizá-los e unificá-los. Os tratamentos foram feitos em uma base PostgreSQL, e a extração dos dados diretamente na ferramenta DBeaver com a construção de um *select* sobre tabelas e *views* para cada alimento (feijões, batatas, arroz e tomate) e tipo de alimento (atacados e varejos), somando-os em oito extrações.

Os parâmetros do método Holt-Winters foram o multiplicativo na tendência e na sazonalidade. Para o período sazonal utilizou-se o valor correspondente ao ciclo anual, onde ocorreu uma variação no valor de cada alimento devido aos seus conjuntos de dados. Após encontrado o número de *clusters*, agrupou-se dois conjuntos de dados para sua aplicação. O primeiro referiu-se aos dados históricos do clima, cotação do dólar, cotação barril do petróleo e preço dos alimentos. E o segundo tratou da junção do resultado obtido ao final da aplicação do Holt-Winters com dados climáticos. Com o conjunto de dados pronto, aplicou-se o K-Means, gerando um peso para cada registro com base na quantidade de clusters, sendo cinco clusters, com pesos de zero a quatro. Estes pesos foram nomeados como "K-classes". Para descobrir a tendência de alta ou baixa dos preços dos alimentos, extraiu-se um conjunto do resultado do K-Means, que era a previsão do Holt-Winters, mas desta vez contendo sua K-classes. Neste conjunto de dados, comparou-se cada registro com o intervalo de preços do seu K-classes pertencente.

4 Resultados

Quanto ao arroz, nas comparações dos resultados do Holt-Winters do atacado obteve-se 40,11% de acertos usando o desvio padrão do conjunto de dados de testes, e para o varejo a comparação com o desvio padrão dos dados de testes foi 48,36%. As comparações utilizando o desvio padrão do histórico ficaram em 100% para os dois tipos de venda. Foram realizadas comparações dos resultados das tendências com os valores reais para o mesmo período, para o atacado foram 8 acertos em 11, e para o varejo 4 acertos em 11 resultados.

Para o feijão preto, as comparações dos resultados do Holt-Winters no atacado tiveram 79,51% de acertos usando o desvio padrão do conjunto de dados de testes, e usando o desvio padrão do histórico foi 87,71%. Para o varejo a comparação com o desvio padrão dos dados de testes foi 10% e para o desvio padrão do histórico foi 37,5%.

Os resultados do feijão cores para o atacado foi 55,84% usando o desvio padrão de testes e com o desvio padrão do histórico foi 92,5%, enquanto para o varejo as porcentagens

foram baixas, 9,02% para o desvio padrão de testes e 18,04% para o desvio padrão do histórico.

Nas análises do K-Means, comparou-se tendências com os valores reais, assim o feijão preto no atacado teve 7 acertos em 11 e no varejo 11 acertos em 11. Para o feijão de cores foram 3 acertos em 11 no atacado e 7 acertos em 11 no varejo.

Quanto a batata inglesa, no Holt-Winters do atacado obteve-se 45% de acertos usando o desvio padrão do conjunto de dados de testes, já para o varejo a comparação com o desvio padrão dos dados de testes foi 64,23%. As comparações usando o desvio padrão do histórico ficaram 80% para o atacado e 66,66% para o varejo. Nas análises do K-Means, comparando os resultados das tendências com os valores reais para o mesmo período, no atacado foram 8 acertos em 11 resultados, e no varejo foram 4 acertos em 11.

Para o tomate, comparando os resultados do Holt-Winters do atacado obteve-se 54,39% de acertos usando o desvio padrão do conjunto de dados de testes, e para o varejo tal comparação obteve 62,5%. Neste caso não foram realizadas comparações com o desvio padrão histórico, pois era menor que o de teste. Para as análises do K-Means comparou-se as tendências com os valores reais e para o atacado a porcentagem de acertos dos resultados analisados foi 6 entre 11 resultado, e para o varejo foi de 5 acertos também entre 11.

5 Conclusões

Os resultados do Holt-Winters tiveram uma porcentagem de acerto de 68,35%, sendo as previsões dos alimentos para atacado e varejo com as respectivas porcentagens do tomate de 54,39% e 62,5%, batata inglesa de 64,23% e 66,66%, feijão preto de 87,71% e 37,5%, feijão cores 92,5% e 18,04% e o arroz com 100% para os dois tipos de venda.

Os resultados gerados pelo K-Means tiveram a porcentagem total de acertos de 57,28%, sendo de atacado e varejo para o alimento tomate de 54,55% e 45,46%, batata inglesa 72,73% e 36,37%, feijão preto 63,64% e 100%, feijão cores 27,28% e 63,64% e o arroz com 72,73% e 36,37%.

REFERÊNCIAS

- [1] Hullen, A. C. N.; Brotto, M. E. *Pobreza e desigualdade no Brasil: para além das políticas de redistribuição, a necessidade das políticas de reconhecimento como forma de combate à pobreza e desigualdade de gênero*, Rio de Janeiro, RJ: Pontifícia Universidade Católica, 2013.
- [2] EMBRAPA. Empresa Brasileira de Pesquisa Agropecuária Ministério da Agricultura, Pecuária e Abastecimento. Mudança do clima e os impactos na agricultura familiar no Norte e Nordeste do Brasil. Disponível em: <https://www.embrapa.br/busca-de-publicacoes/-/publicacao/1046425/mudanca-do-clima-e-os-impactos-na-agricultura-familiar-no-norte-e-nordeste-do-brasil>, Acesso em 25 de outubro de 2020.
- [3] Suppalakpanya, K.; Nikhom, R.; Booranawong, A.; Booranawong, T. An evaluation of holt-winters methods with different initial trend values for forecasting crude palm oil production and prices in thailand - Suranaree J. Sci. Technol. v. 26 n. 1; jan./mar. 2019.
- [4] Trata, L. F.; Strmcnik E. *The comparison of HolteWinters method and Multiple regression method: A case study*. v. 109 p. 266-276; apr. 2016.