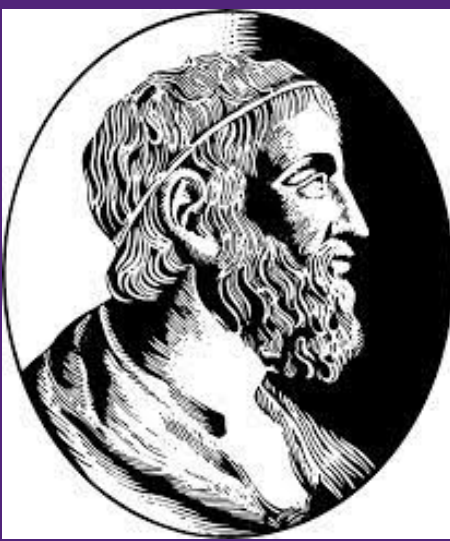


Previsão do Consumo de Cimento no Brasil

Julia Leite ¹ Prof. Dr. Marcelo Finger

¹Instituto de Matemática e Estatística - Universidade de São Paulo



Introdução

A falta de um modo bem fundamentado para prever a demanda de cimento é uma dor entre as empresas cimenteiras, uma vez que tanto a construção de uma fábrica quanto a ampliação na capacidade de produção de uma existente são custosas e demoradas. Dessa forma, um modelo que permitisse prever o consumo a nível de estados do Brasil poderia auxiliar gestores na tomada de decisões e apoiar órgãos governamentais a direcionar ações para mitigar o grande impacto ambiental da fabricação desse produto.

Objetivos

Este trabalho, então, propõe-se a aplicar modelos de aprendizado de máquina para determinar qual é mais eficiente para prever a demanda por cimento nos estados do Brasil. Os modelos avaliados são: regressão linear, redes neurais *multi-layer perceptron* (MLP) e redes neurais recorrentes.

Dados

Os modelos recebem como entrada dados econômicos de um estado e em um mês específico e têm como resposta (*output*) o consumo de cimento no mês seguinte nesse estado. A exceção são as redes neurais recorrentes, pois recebem dados de meses anteriores ou de outros estados.

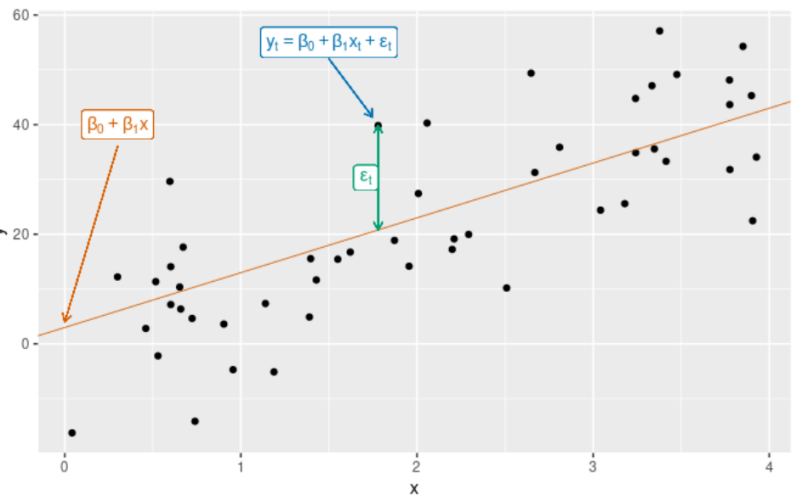
| Categoria | Indicadores |
|---------------------------------|---|
| Crescimento econômico | PIB do estado, PIB <i>per capita</i> , população , PIB da construção civil e desemprego |
| Política Fiscal | Necessidade de Financiamento do Setor Público (NFSP) e Estoque da Dívida Pública |
| Inflação | Índice de Preços ao Consumidor Aplicado (IPCA), Índice Nacional de Custo da Construção (INCC) e Índice Geral de Preço (IGP) |
| Política monetária | Taxa SELIC |
| Indicadores sociais | Índice de Desenvolvimento Humano (IDH) |
| Indicadores da Construção Civil | Produção de cimento, preço do saco de 50kg, da tonelada e do quilograma de cimento |

Figura 1. Dados de entrada (indicadores) utilizados

Realizou-se uma análise exploratória dos dados afim de direcionar a estratégia de processamento, preparação e limpeza de dados. A partir dessa análise, optou-se por utilizar dados de 2003 até 2019, com granularidade mensal, para realizar o estudo.

Modelos

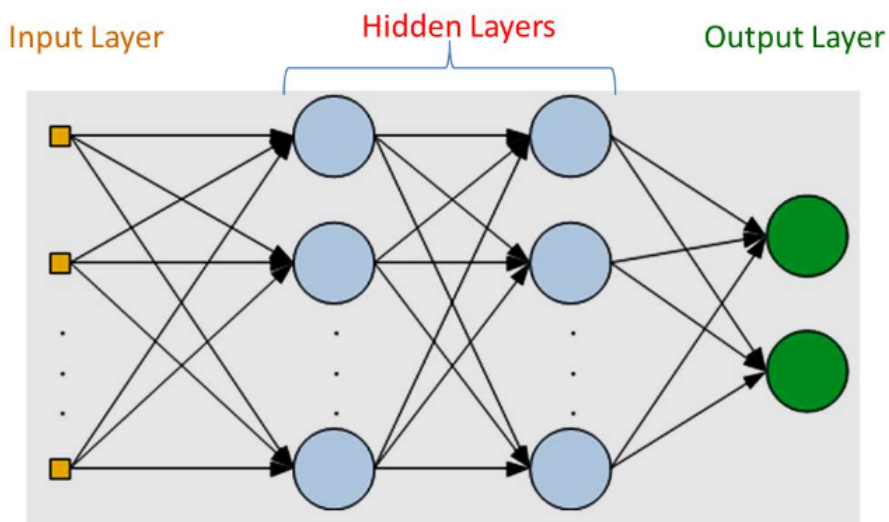
Regressão linear



O modelo de regressão linear supõe uma relação linear entre a variável que será prevista (*target*) e as variáveis de entrada (*predictor*). Calcula, então, um coeficiente para cada *predictor* afim de mensurar o efeito desse no *target*, de modo a minimizar o erro na previsão.[1]

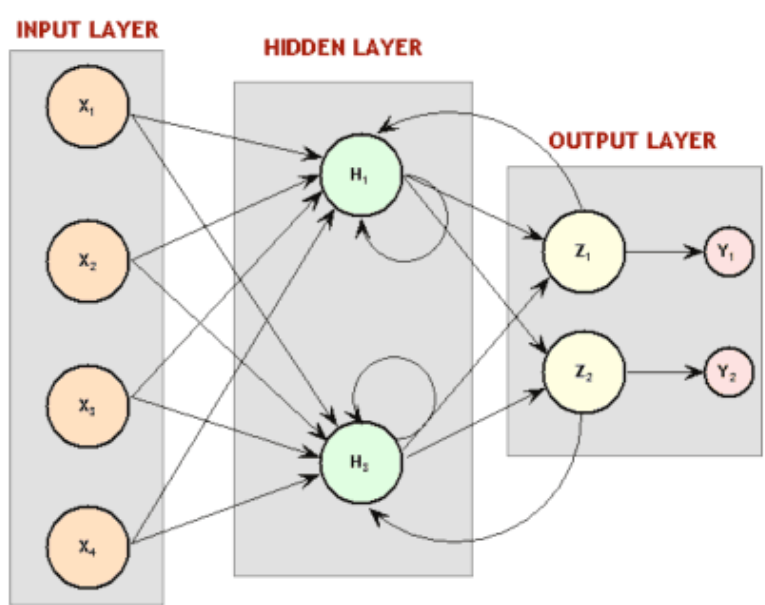
$$y_t = \beta_0 + \beta_1 x_{t_1} + \beta_1 x_{t_2} + \dots + \beta_k x_{t_k} + \varepsilon$$

Redes neurais *multi-layer perceptrons* (MLPs)



Redes neurais são modelos de *machine learning* inspiradas no cérebro humano aonde o aprendizado ocorre ao combinar neurônios matemáticos que estabelecem conexões de acordo com o treinamento fornecido. Neste trabalho, aplicaram-se redes de arquitetura *multilayer perceptrons* (MLPs), ou seja, com múltiplas camadas de neurônios. [2]

Redes Neurais Recorrentes



Já as redes neurais recorrentes são projetadas para reconhecer padrões nos dados, uma vez que levam tempo e sequência em consideração. Assim, nessas redes, a decisão tomada na etapa anterior influencia a etapa seguinte por conta dos *loops de feedback*, então o presente e o passado recente se combinam para determinar a previsão. Neste trabalho, foram testadas as redes Long Short Term Memory (LSTM), redes Gated Recurrent Unit (GRU) e Bidirecionais.[2]

Avaliação de performance

A performance dos modelos foi avaliada por meio de três métricas estatísticas: a média dos erros absolutos (MAE), do inglês *mean squared error*, a *root mean squared error* (RMSE) e a *mean absolute percentage error* (MAPE). Além disso, calculou-se o delta percentual (Δ) de cada previsão afim de visualizar a distribuição do erro percentual, ou seja, se os modelos tendem a superestimar ou subestimar o consumo. Abaixo, as fórmulas matemáticas das métricas citadas:

$$MAE = \frac{1}{n} \sum_{j=1}^n |y_j - \hat{y}_j| \quad RMSE = \sqrt{\frac{1}{n} \sum_{j=1}^n (y_j - \hat{y}_j)^2} \quad MAPE = \frac{1}{n} \sum_{t=1}^n \left| \frac{y_t - \hat{y}_t}{y_t} \right| \quad \Delta = \frac{|\hat{y}_i - y_i|}{\hat{y}_i} \quad (1)$$

Resultados

Para cada um dos modelos, foram realizados testes para determinar a combinação de parâmetros que resultava em um melhor desempenho. Por exemplo, no caso da regressão linear, os melhores resultados foram obtidos ao remover variáveis de entrada com alto índice de correlação. Abaixo, se encontra a avaliação do desempenho de cada um dos modelos utilizados no estudo:

| Modelo | MAPE | MAE | RMSE |
|---------------------------|------|-----------|-----------|
| Regressão linear | 33% | 34095 ton | 58242 ton |
| Redes neurais MLP | 26% | 22538 ton | 36371 ton |
| Redes neurais recorrentes | 18% | 19845 ton | 34513 ton |

Interessante notar que, apesar de os valores de erro médio absoluto e quadrado (MAE e RMSE) parecerem altos, o consumo médio de cimento no período de julho de 2017 a dezembro de 2019 foi de 147 176 ao mês.

Observa-se que a rede neural recorrente apresenta melhor performance frente à regressão linear e à rede MLP.

Previsão

No gráfico abaixo, se encontra a distribuição dos deltas percentuais das previsões realizadas com as redes neurais recorrentes.

Distribuição das faixas de erro

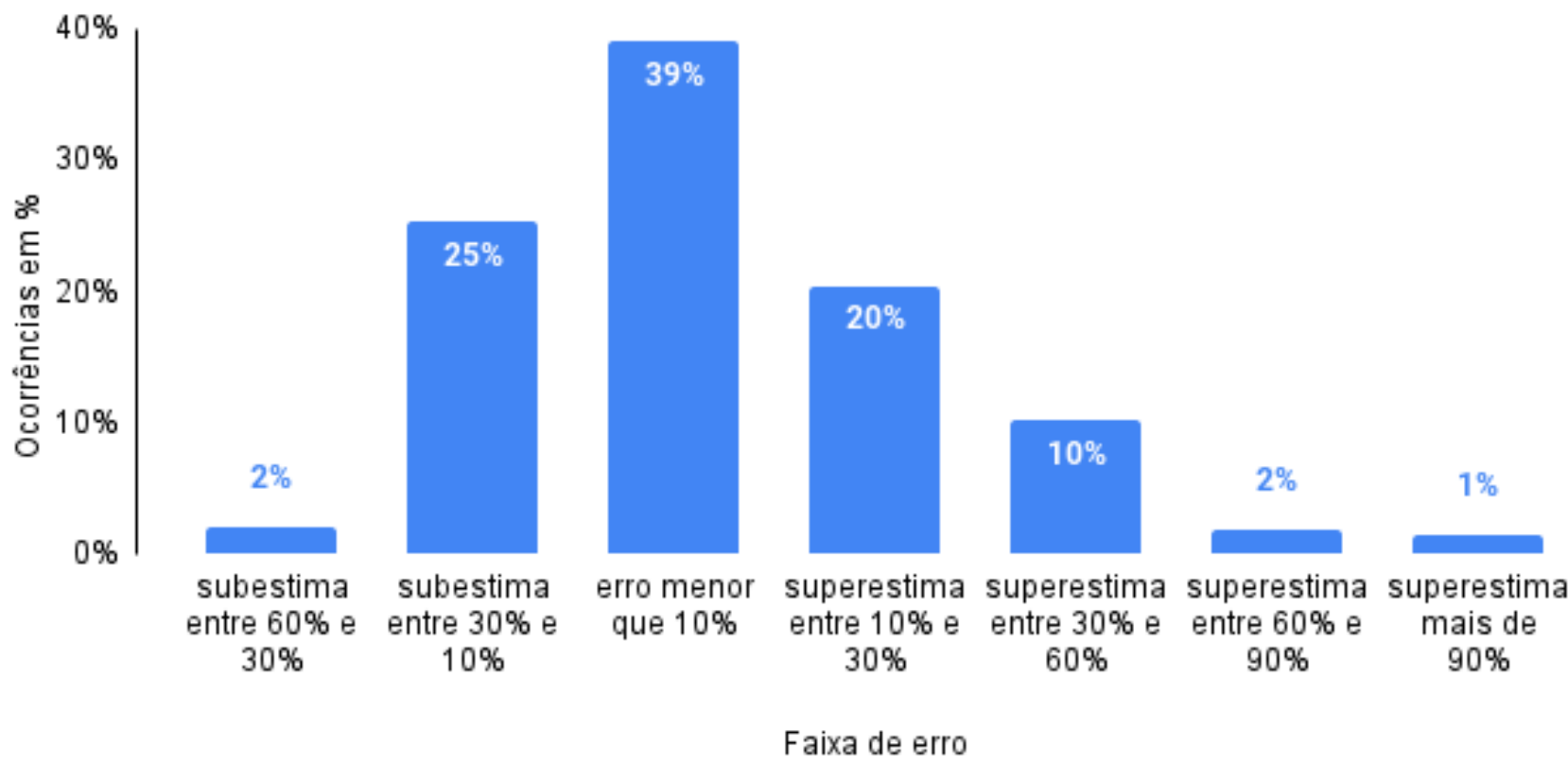


Figura 2. Distribuição dos deltas percentuais

O gráfico indica uma significativa taxa de acerto do modelo, já que 84% das previsões apresentaram uma taxa de erro absoluta inferior a 30%.

Foram selecionados os estados de São Paulo (SP), Minas Gerais (MG), Ceará (CE) e Paraná (PR) para visualizar o comportamento do modelo. No gráfico abaixo, cada linha contínua representa a previsão realizada pelo modelo para um estado em específico, a linha pontilhada, por sua vez, corresponde à real demanda por cimento no estado correspondente.

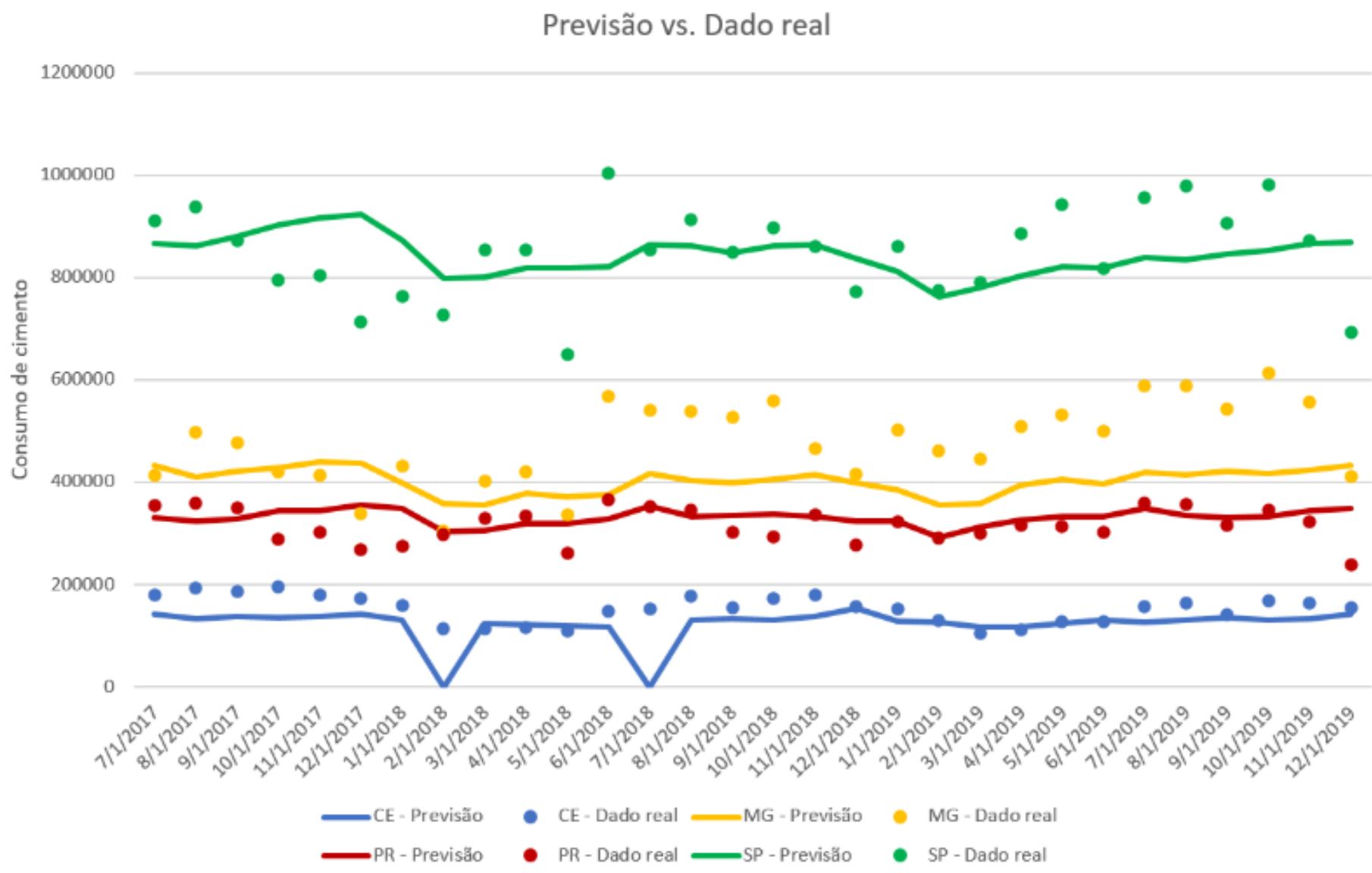


Figura 3. Previsão feita pelo modelo em comparação com o consumo real

Pode-se notar que a previsão realizada pelo modelo não coincide perfeitamente com os dados de consumo, comportamento esperado uma vez que é comum que haja erro de medição, ruídos ou comportamentos anormais nos dados. Contudo, as previsões fornecidas pelo modelo são capazes de seguir a tendência apresentada nos dados.

Conclusão

A partir da análise do desempenho dos três modelos estudados, observou-se o ganho ao utilizar as redes neurais recorrentes, uma vez que, por levarem em consideração informação histórica, apresentam melhor performance frente à regressão linear e às redes MLP. Conclui-se, ainda, que os modelos de *machine learning* apresentam bom potencial e podem ser utilizados para prever a demanda por cimento no Brasil.

Referências

- [1] Hyndman, R.J., Athanasopoulos, G. (2021) Forecasting: principles and practice, 3rd edition, OTexts: Melbourne, Australia. OTexts.com/fpp3. Accessed on 09 December 2022.
- [2] Ian Goodfellow, Yoshua Bengio e Aaron Courville. *Deep Learning*. [http : / / www . deeplearningbook.org](http://www.deeplearningbook.org). MIT Press, 2016.