

UNIVERSIDADE DE SÃO PAULO
INSTITUTO DE MATEMÁTICA E ESTATÍSTICA
BACHARELADO EM CIÊNCIA DA COMPUTAÇÃO

**Previsão de consumo de cimento através
de modelos de Machine Learning**

Julia Leite da Silva

MONOGRAFIA FINAL

MAC 499 — TRABALHO DE
FORMATURA SUPERVISIONADO

Supervisor: Prof. Dr. Marcelo Finger

São Paulo
2022

*O conteúdo deste trabalho é publicado sob a licença CC BY 4.0
(Creative Commons Attribution 4.0 International License)*

Dedico este trabalho a minha família, meu pai e minha mãe que têm iluminado meu dia e me dado força, alegria e paz. Acima de tudo, dedico a Deus, quem me presentou com capacidade, com uma família maravilhosa e que foi adicionando pessoas no meu caminho para me apoiar e me ajudar em cada passo da jornada.

Agradecimentos

Fight with determination, embrace life and live it with passion. Lose your battles with class and dare to win because the world belongs to those who dare to live. Life is worth too much to be insignificant.

— Charles Chaplin

Queria agradecer, antes de tudo, a Deus por tudo que Ele tem me dado. Uma vez minha mãe me disse que Deus abençoa o trabalho das nossas mãos, que várias vezes onde ela pensou que não daria certo ou não daria conta, ela pedia ajuda e, no final, tudo se endireitava. Ela me disse também que Deus colocava pessoas na vida dela que a ajudavam e orientavam a encontrar um caminho. Fico muito grata de enxergar isso acontecendo na minha vida também.

Agradeço aos meus pais, que têm sido minha força, alegria e lar. Agradeços aos meus amigos e amigas, por colorir meus dias e encher meu rosto de sorrisos.

Quero também agradecer ao Marcelo Finger, meu orientador obrigada para o Felipe, que me surpreendeu com a parceria e a nessa jornada. Muito ajuda. Agradecer também ao Gabriel, pelo apoio para escrever esse trabalho.

Resumo

Julia Leite da Silva. **Previsão de consumo de cimento através de modelos de Machine Learning**. Monografia (Bacharelado). Instituto de Matemática e Estatística, Universidade de São Paulo, São Paulo, 2022.

[illegible]

Palavras-chave: Palavra-chave1. Palavra-chave2. Palavra-chave3.

Abstract

Julia Leite da Silva. . Capstone Project Report (Bachelor). Institute of Mathematics and Statistics, University of São Paulo, São Paulo, 2022.

[illegible]

Keywords: Keyword1. Keyword2. Keyword3.

Lista de Abreviaturas

| | |
|------|--|
| MLP | Perceptron Multi-Camadas (<i>Multilayer perceptron</i>) |
| LSTM | (<i>Long Short Term Memory</i>) |
| RMSE | Raíz Média de erros Quadrados (<i>Root Mean Squared Error</i>) |
| MAPE | Erro Percentual Absoluto Médio (<i>Mean Absolute Percentage Error</i>) |

Lista de Símbolos

| | |
|----------|-----------------------------------|
| ω | Frequência angular |
| ψ | Função de análise <i>wavelet</i> |
| Ψ | Transformada de Fourier de ψ |

Lista de figuras

Lista de tabelas

Lista de programas

Sumário

| | |
|--------------------------------|---------------|
| Introdução | 1 |
| 1 Fundamentação teórica | 3 |
| 2 Metodologia | 5 |
| 3 Experimentos | 7 |
| 3.1 regressao linear | 7 |
| 4 Resultados | 9 |
| 5 Conclusão | 11 |
| Referências | 13 |

Introdução

O que prédios, pontes, hidrelétricas e aeroportos têm em comum? Todos são frutos da indústria da construção civil, um importante componente do investimento brasileiro e, conseqüentemente, uma das grandes engrenagens responsáveis por movimentar a atividade econômica no Brasil. Em 2021, por exemplo, o Produto Interno Bruto (PIB) desse setor registrou alta de 9,7%, enquanto o PIB do Brasil cresceu 4,6%, assim, o setor da construção figura como importante impulsionador da economia do país. [VASCONCELOS, 2022](#) O cimento, nesse contexto, por ser um ingrediente central da argamassa e do concreto, caracteriza-se como um dos principais insumos da indústria.

Motivação

Contudo, a falta de um modo bem fundamentado para prever a demanda de cimento é uma dor entre as empresas cimenteiras, uma vez que a construção de uma fábrica é custosa e demorada, além disso, aumentar a capacidade de produção de uma fábrica também é um processo custoso.

Dessa forma, um modelo que permitisse prever a demanda a nível de estados do Brasil poderia auxiliar gestores a tomar melhores decisões e a estruturar a estratégia de forma mais embasada, de modo a reduzir os riscos do setor. Além disso, poder-se-ia apoiar órgãos governamentais a direcionar ações para mitigar o 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.

Trabalhos correlatos

Capítulo 1

Fundamentação teórica

Problema

Os problemas de *machine learning* pertencem a duas categorias principais: classificação ou regressão. As tarefas de classificação, por um lado, consistem em determinar a categoria, dentre as k disponíveis, a que um *input* pertence, um exemplo de tarefa de classificação seria determinar se o consumo de cimento em um estado em um mês específico representa um aumento, queda ou estabilidade em relação ao mês anterior. Os problemas de regressão, por outro lado, compreendem prever um valor numérico a partir de um *input*, [GOODFELLOW et al., 2016](#) como adotado neste trabalho.

Regressão linear

O modelo de regressão linear estabelece uma relação linear, ou seja, uma função, entre a variável que será prevista (*target*) e as variáveis de entrada (*predictor variables*), neste projeto: o consumo de cimento em um mês e estado determinados e os indicadores econômicos correspondentes, respectivamente. O algoritmo, então, calcula um coeficiente para cada *predictor* afim de mensurar o efeito desse no *target*, de modo a minimizar o erro na previsão.

Redes neurais

Redes neurais multi-layer perceptrons

Redes neurais são modelos de *machine learning* inspiradas no cérebro humano aonde o aprendizado ocorre ao se agregarem neurônios matemáticos que estabelecem conexões de acordo com o treinamento fornecido. Neste trabalho, aplicaram-se redes *multilayer perceptrons* (MLPs), ou seja, que apresentam múltiplas camadas de neurônios e *feedforward*, onde a saída de uma camada de neurônios é utilizada como entrada para a camada seguinte, sem utilizar retropropagação.

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.

GRU

LSTM

Função de ativação

Normalização

Capítulo 2

Metodologia

Os modelos recebem como entrada dados econômicos obtidos do Instituto Brasileiro de Geografia e Estatística (IBGE), Fundação Getúlio Vargas (FGV), Instituto de Pesquisa Econômica Aplicada, entre outros. Sobre crescimento econômico, foram utilizados: PIB do estado, PIB per capita, população, PIB da construção civil e desemprego. Já para política fiscal, utilizou-se: Necessidade de Financiamento do Setor Público (NFSP) e Estoque da Dívida Pública. Para mensurar a inflação foram utilizados Índice de Preços ao Consumidor Aplicado (IPCA), Índice Nacional de Custo da Construção (INCC) e Índice Geral de Preço (IGP). O Índice de Desenvolvimento Humano (IDH) foi utilizado como indicador social. Finalmente, a produção de cimento e o preço do saco de 50kg, da tonelada e do quilograma de cimento foram utilizados como indicadores da construção civil.

A partir da análise exploratória dos dados de entrada, optou-se por utilizar dados de 2003 até 2019, com granularidade mensal, para realizar o estudo. Se os indicadores apresentavam granularidade anual, calculou-se a divisão da medição de cada ano por 12 meses para obter a média mensal. Além disso, caso o dado estivesse disponível apenas a nível de Brasil, ao invés de por estado, o valor da medição da União foi utilizado para todos os estados.

A estratégia utilizada para lidar com dados faltantes nas variáveis de entrada foi repetir o valor da ocorrência anterior ou marcar a entrada com um valor não presente no intervalo de dados, a exemplo de marcar -1"como valor da produção mensal de cimento em um determinado mês em um estado específico.

Finalmente, os dados utilizados como entrada foram deslocados um mês à frente ou um ano, no caso dos indicadores anuais, em relação aos dados de consumo. Dessa forma, os dados correspondentes a, por exemplo, fevereiro de 2004 estão relacionados ao consumo de cimento em março de 2004, com o objetivo de propor um cenário mais pertinente, uma vez que o objetivo do projeto é prever a demanda por cimento no mês seguinte em um estado a partir dos dados do mês atual e, eventualmente, dos anteriores.

Capítulo 3

Experimentos

3.1 regressao linear

Capítulo 4

Resultados

blafgdf .

Capítulo 5

Conclusão

ahahahahahh

Referências

- [GOODFELLOW *et al.* 2016] Ian GOODFELLOW, Yoshua BENGIO e Aaron COURVILLE. *Deep Learning*. <http://www.deeplearningbook.org>. MIT Press, 2016 (citado na pg. 3).
- [VASCONCELOS 2022] Ieda VASCONCELOS. *Informativo Econômico PIB*. 2022. URL: <https://cbic.org.br/wp-content/uploads/2022/03/informativo-economico-pib-04-marco-2022.pdf> (citado na pg. 1).