
Título: Taxa SELIC

Thomas Kenzo Matsuta e Edilberto Silva

FACSENAC-DF, Brasília-DF

kenzomatsuta@gmail.com; prof.edilberto.silva@gmail.com

Abstract: *Selic is the economy's basic interest rate. It is the main monetary policy instrument used by the Central Bank (BC) to control inflation. It influences all interest rates in the country, such as interest rates on loans, financing and financial investments.*

The name of the Selic tax comes from the acronym of the Special System of Settlement and Custody. Such a system is a financial market infrastructure managed by the Central Bank. It trades federal public bonds. The adjusted average rate of daily financing calculated in this system corresponds to the Selic rate. This article presents the data and the results of the Selic rate per month obtained in the participation of this demand for the elaboration of models of classification algorithm, for this purpose the CRISP-DM methodology was applied, as it is specific for data mining processes. The Word2vec method was also used with the Skip-gram algorithm to generate the data dictionary and the embeddings to generate contexts.

Resumo. A Selic é a taxa básica de juros da economia. É o principal instrumento de política monetária utilizado pelo Banco Central (BC) para controlar a inflação. Ela influencia todas as taxas de juros do país, como as taxas de juros dos empréstimos, dos financiamentos e das aplicações financeiras.

O nome da taxa Selic vem da sigla do Sistema Especial de Liquidação e de Custódia. Tal sistema é uma infraestrutura do mercado financeiro administrada pelo BC. Nele são transacionados títulos públicos federais. A taxa média ajustada dos financiamentos diários apurados nesse sistema corresponde à taxa Selic. Neste artigo são apresentadas as datas e os resultados da taxa selic ao mês obtidos na participação dessa demanda de elaboração de modelos de algoritmo de classificação, para tal foi adotada a metodologia CRISP-DM por ser específica para processos de mineração de dados. Foi utilizado ainda o método Word2vec com o algoritmo Skip-gram para geração do dicionário de dados e os embeddings para geração de contextos.

Palavras-chave: economia, taxa selic, juros, GBM, word2vec.

1. Introdução

O presente trabalho tem como foco a taxa selic já que ela é uma referência para o custo das linhas de crédito em geral. Quando ela é elevada, a tendência é de que empréstimos e financiamentos fiquem mais caros.

Nesse contexto, o objetivo deste trabalho é realizar um Business Intelligence, para comparar o poder de compra dos cidadãos no DF entre 2020 e 2023.

Para cumprir com o objetivo foram definidos os seguintes objetivos específicos (figura 2):

- **Análise e estudo algoritmos e técnicas de mineração de textos com fulcro a classificá-los;**
- **Obtenção e tratamento da massa de dados necessária para responder o objetivo geral**
- **Identificar o cumprimento do padrão de qualidade esperado, seguindo a seguinte regra: o atendente terá que ter postado necessariamente ao menos uma mensagem de saudação, uma de esclarecimento ou oferta de produtos e serviços e uma mensagem de encerramento.**
- **Análise, treinamento dos algoritmos necessários para encontrar padrões com vistas a responder a pergunta-chave**
- **Validação e ajuste dos padrões obtidos junto a área demandante/cliente.**

Nos dias de hoje, o BI, tem se mostrado de grande importância, aumentando assim a competitividade e o impulso diante dos resultados empresariais. É uma solução de grande influência, ficando claro o ganho na velocidade das respostas dentro da instituição tanto no nível estratégico, tático e operacional. Composto por profissionais, infraestrutura, ferramentas e dados, a união de todos esses componentes é o que permite a realização de um sistema de apoio à decisão dentro de uma instituição.

O processo de BI inicia-se com a coleta de dados, limpeza e a transformação dos dados, o chamado ETL (Extract, Transform, Load) que são um conjunto de ferramentas que fazem todo esse processo e ao final carregam no Data Warehouse (Armazém de Dados)

que tem a função de associar os dados coletados, criando um repositório central ou depósito de dados.

2. Trabalhos relacionados e referência teórica

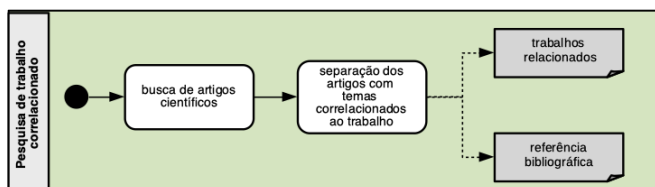


Figura 2: Fluxo de atividades - Pesquisa de trabalhos relacionados

A Selic é a taxa básica de juros da economia. É o principal instrumento de política monetária utilizado pelo Banco Central (BC) para controlar a inflação. Ela influencia todas as taxas de juros do país, como as taxas de juros dos empréstimos, dos financiamentos e das aplicações financeiras.

O nome da taxa Selic vem da sigla do Sistema Especial de Liquidação e de Custódia. Tal sistema é uma infraestrutura do mercado financeiro administrada pelo BC. Nele são transacionados títulos públicos federais. A taxa média ajustada dos financiamentos diários apurados nesse sistema corresponde à taxa Selic.

Quando ela é elevada, a tendência é de que empréstimos e financiamentos fiquem mais caros. Crédito e consumo andam lado a lado. Quando os empréstimos e financiamentos ficam mais caros, naturalmente o nível de consumo tende a diminuir.

O Departamento Intersindical de Estatística e Estudos Socioeconômicos (DIEESE) é uma entidade criada e mantida pelo movimento sindical brasileiro. Foi fundado em 1955, com o objetivo de desenvolver pesquisas que subsidiam as demandas dos trabalhadores. Sindicatos, federações, confederações de trabalhadores e centrais sindicais são filiados ao DIEESE e fazem parte da direção da entidade. Atualmente, são cerca de 700 associados.

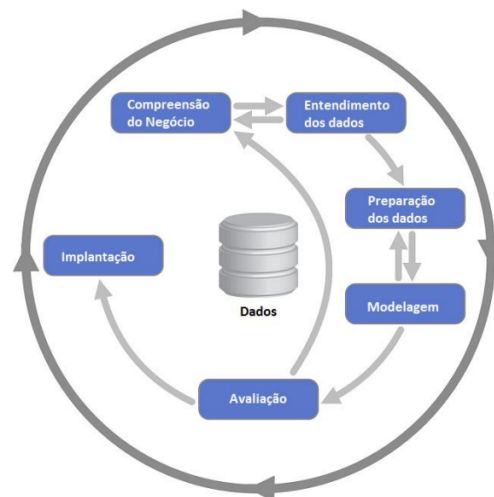


Figura 2 - Fases do CRISP-DM. Fonte: [Shearer, 2000]

Durante o processo desse trabalho, foram realizadas pesquisas em artigos com temas relacionados, com intuito de ter uma maior compreensão a respeito dos conceitos e a da criação de ambientes de Business Intelligence. Um dos trabalhos analisados teve como objetivo principal escolher uma plataforma de BI Open Source² que se contribui com a melhoria da gestão da informação no ambiente organizacional. Por meio de pesquisa foi escolhido o SpagoBI, utilizado para elaboração de um estudo de caso que mostra como as técnicas de BI podem apoiar a análise de dados de órgão da administração pública. Foi realizada uma análise com dados da CGU por meio do portal da Transparência em formato CSV, dados de servidores da administração pública, penalizados nos últimos anos.

Em uma segunda pesquisa analisada, o objetivo do trabalho foi a analisar a aplicação de sistema de BI no suporte à decisão, de dados operacionais de aplicativos no Governo Eletrônico [4], definido como a produção e disponibilização de serviços governamentais através das TIC, composta por ferramentas e tecnologias de código aberto. Foi utilizado código aberto para possibilitar a criação de um ambiente de BI onde não existem custos com licenças de software, sendo possível encontrar ferramentas com documentação disponível para acesso gratuito e com uma comunidade ativa trabalhando na evolução destas ferramentas. [3]

O objetivo proposto do trabalho foi alcançado, usando ambientes de BI como solução para análise e interpretação dos dados, utilizando códigos abertos e realizando todo o processo de ETL até o usuário final. Foi apresentado um estudo de caso, onde o ambiente de BI proposto foi criado utilizando a metodologia descrita no trabalho. Aplicando metodologias e ferramentas para governança das TIC e dados do

ambiente de BI, e a análise de outros atributos de qualidade do ambiente, como disponibilidade e performance das consultas aos dados e acessibilidade.

3. Metodologia e CRISP-DM

Esta pesquisa foi descritiva de natureza quantitativa onde analisou-se uma base de dados existente sobre valores da taxa Selic e das cestas básicas no DF entre 2020 e 2023. Neste trabalho foi utilizada a metodologia CRISPDM (Cross-Industry Standard Process for Data Mining), que tem como objetivo orientar a condução de um projeto de mineração de dados. Tem como passos a compreensão do negócio; o entendimento dos dados com a inspeção, organização e descrição dos atributos para se ter o entendimento de cada um e qual sua relevância; na preparação utiliza-se as informações obtidas e gera-se um conjunto de dados (dataset) pronto para ser usado nas fases seguintes; na modelagem aplica-se um modelo de classificação; na avaliação o confronto do modelo gerado com novos dados e a análise do resultado; por fim, conclui-se com a apresentação dos parâmetros utilizados no algoritmo escolhido e resultados alcançados.

3.1. Compreensão do Negócio

O foco deste trabalho é voltado para a identificação dos valores da cesta básica e comparar com os valores da taxa Selic vindo do BACEN (Banco Central) e DIEESE (Departamento Intersindical de Estatística e Estudos Socioeconômicos).

Dentre os muitos, e possíveis, critérios que podem ser utilizados no processo de avaliação de qualidade dos atendimentos, para esse trabalho, o escopo da qualidade definido foi:

1. Classificação dos valores da taxa Selic e do preço das cestas básicas do local e data especificada.
2. Comparar de lado a lado e chegar a uma conclusão.

A Selic é a taxa básica de juros da economia. É o principal instrumento de política monetária utilizado pelo Banco Central (BC) para controlar a inflação. Ela influencia todas as taxas de juros do país, como as taxas de juros dos empréstimos, dos financiamentos e das aplicações financeiras. O nome da taxa Selic vem

da sigla do Sistema Especial de Liquidação e de Custódia.

A análise deste estudo foi realizada com base nos dados do Banco Central e do DIEESE (Departamento Intersindical de Estatística e Estudos Socioeconômicos) que é o órgão responsável pela regulação dos preços das cestas básicas. Foi fundada em 1955 para desenvolver pesquisas que sirvam de apoio às reivindicações dos trabalhadores, para ser um ponto de partida nas suas questões trabalhistas. A maioria das confederações e federações nacionais de trabalhadores, bem como os principais sindicatos, estão filiados no DIEESE, estando na sua liderança e contribuindo para o seu sustento.

O DIEESE é um órgão formado por uma equipe multidisciplinar de economistas, sociólogos, estatísticos, geógrafos e outros profissionais que fornecem produtos exclusivos e serviços diferenciados aos sindicatos. As questões estruturais estudadas e pesquisadas.

Desde a sua criação, o DIEESE realiza pesquisas e indicadores permanentes com o objetivo de fornecer informações sobre renda, custo de vida e mercado de trabalho e contribuir com os trabalhadores para melhorar o conhecimento e a compreensão da estrutura da realidade.

Cesta básica é um conjunto formado por produtos utilizados por uma família durante um mês.

Existem diversos modelos de cesta básica, não havendo atualmente uma formulação precisa de quais produtos a compõem.

A noção de um conjunto mínimo de produtos alimentícios necessários para a nutrição de uma família durante um mês possui suas origens ligadas às definições de salário mínimo; até hoje, uma série de cestas básicas foram e continuam a ser propostas.

No Brasil, apesar de existirem diversos padrões de cestas básicas, aquele estabelecido pelo DIEESE é o mais comum. A cesta básica elaborada pelo DIEESE adquire preços e conteúdos diferentes de acordo com a região onde é montada. No geral, todas as cestas básicas são montadas com quantidades variadas de 13 gêneros alimentícios considerados necessários para a manutenção da vida: carne, leite, feijão, arroz, farinha, batata, tomate, pão, café em pó, frutas, açúcar, óleo (gordura) e manteiga.

3.2. Entendimento de Dados

Tal como descrito na figura 3, foram extraídos de bases corporativas do BACEN e DIEESE, arquivos (Apêndice A) em formato .csv, os quais foram utilizados para criar os embeddings, os quais foram utilizados na realização do treinamento do algoritmo e teste do modelo de classificação gerado.

Taxa Selic		Cesta Básica	
data	valor	data	valor
jan/20	4,4	jan/20	483,17
fev/20	4,19	fev/20	481,78

Data | Valor da taxa selic (base 252)

Data | Valor da cesta basica (reais)

Os dados técnicos obtidos pelo BACEN atualizam mensalmente. A taxa Selic está em meses (a.m) e no formato de base 252. Um banco central, banco de reservas ou autoridade monetária é uma entidade independente ou ligada ao Estado cuja função é gerir a política econômica, ou seja, garantir a estabilidade e o poder de compra da moeda de cada país e do sistema financeiro como um todo. Além disso, tem como objetivo definir as políticas monetárias (taxa de juros e câmbio, entre outras) e aquelas que regulamentam o sistema financeiro local. O banco faz isso interferindo mais ou menos no mercado financeiro, vendendo papéis do tesouro, regulando juros e avaliando os riscos econômicos para o país. Os arquivos de Série Histórica de Preços da taxa Selic e da cesta básica - Sua atualização ocorre todos meses, sendo disponibilizado mensalmente ou anualmente, cada arquivo, que se encontram em formato .csv, sendo utilizado 2 arquivo um do DIEESE no link:

<https://www.dieese.org.br/cesta/>

E outro do BACEN no link:

<https://dadosabertos.bcb.gov.br/dataset/1178-taxa-de-juros---selic-anualizada-base-252>

campo	descrição
Mês (data)	Mês da taxa selic
Ano (data)	Ano da taxa selic

Valor (taxa_selic)	Valor da taxa Selic
Mês (data)	Mês da cesta_basica
Ano (data)	Ano da cesta_basica
Valor (cesta_basica)	Valor da cesta_basica

3.3. Preparação de Dados

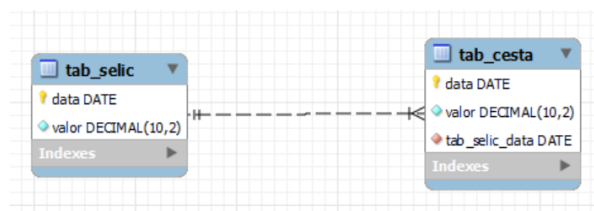
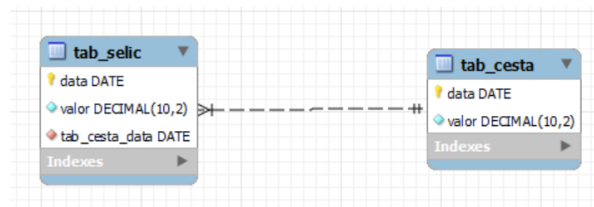
Essa etapa tem o objetivo de tratar os dados antes de serem submetidos ao treinamento, que já foram filtrados e tratados no excel.

3.3.1 Elaboração do Staging Area

Para montar a staging area com todas as informações necessárias foram extraídos dos bancos de dados do BACEN e do DIEESE e manipulados para gerar as respectivas tabelas de carga das fases seguintes.

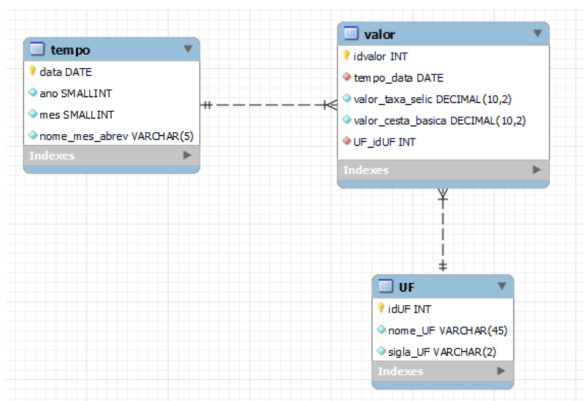
<https://github.com/ItsukaT2/SQLs>

As estruturas das tabelas são compostas dos seguintes dados:



3.3.2 ETL

O processo de tratamento e carga de dados se inicia no segundo fluxo de ETL figura 4 criado. Esse fluxo é referente à dimensão tempo que obtém as informações do arquivo .csv e adiciona a eles os campos ano, mes, nome_do_mes_abrev para completar a dimensão.



3.3.3 Modelo Multidimensional

Após a inserção dos dados na Staging Area....

Na primeira camada, serão extraídos dos bancos de dados dos órgãos os arquivos .csv que será utilizado na etapa de ETL.

Na segunda camada, de ETL, utilizou-se o Excel, para formatar, arrumar e organizar os valores.

Na terceira camada através da ferramenta excel para consultas em tempo real e auxiliando no desenvolvimento do Dashboard.

Na quarta e última camada será utilizado o Pentaho BI Server com algumas de suas ferramentas. Para realizar consultas no cubo de acordo com as necessidades apontadas pelo especialista, foi utilizada uma ferramenta OLAP denominada Saiku. Essas consultas serviram de base para o desenvolvimento de um Dashboard, através do Community Dashboard Editor - CDE.

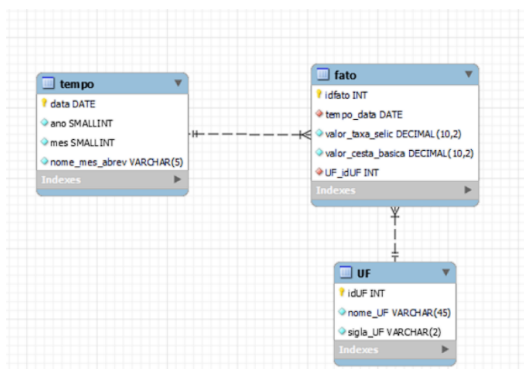


figura 1: modelo dimensional

O modelo multidimensional foi esquematizado e desenvolvido com o auxílio da ferramenta MySQL Workbench. O script SQL pode ser visto em:

<https://github.com/ItsukaT2/SQLs>

Os 02 (dois) arquivos .csv foram colocados em uma pasta chamada Dados; 01 (um) arquivo da Série Histórica de Preços de cestas básicas, e por último o arquivo Valor da taxa selic colocado na mesma pasta.

Na visão multidimensional será atribuído modelo da figura 1, facilitando e dando agilidade nas pesquisas que será realizada na base de dados, contendo apenas um modelo multidimensional Valor da taxa/cesta e tempo tendo em vista os indicadores apresentados no estudo de caso.

3.4. Modelagem

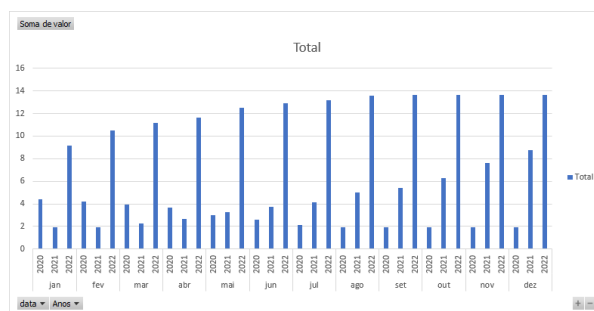


Gráfico: Taxa Selic

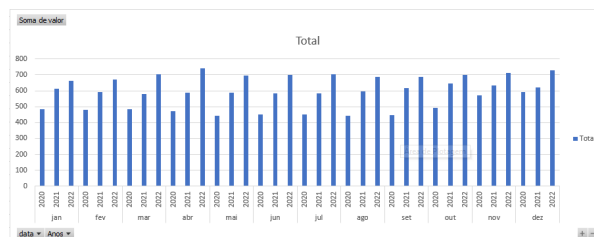


Gráfico: Valor da Cesta Básica

As imagens encontra-se no link:

<https://github.com/ItsukaT2/SQLs>

3.5. Avaliação

Nesta seção são apresentados os resultados extraídos da estrutura de BI criada, mostrando a relação desses resultados com os objetivos específicos descritos no capítulo 1. Conforme mencionado, a

análise realizada se refere às propostas cadastradas nos dados do BACEN e do DIEESE, de janeiro de 2020 até janeiro de 2023.

Em linhas gerais, do ano de 2020 até janeiro de 2023, todos os valores da taxa selic se comportaram de forma semelhante com o valor das cestas básicas.

consumo e do investimento, e levar ao aumento ou diminuição dos preços das cestas básicas.

4. Conclusão

No trabalho desenvolvido foi abordada a relação com o valor da taxa selic com o valor das cestas básicas, já que a selic tem uma alta relação com o IPCA (Índice de preços ao consumidor). Quando a taxa Selic está alta, as taxas de juros de empréstimos e financiamentos também tendem a ser elevadas, tornando o crédito mais caro. Nesse cenário, as pessoas e empresas podem se sentir desencorajadas a fazer empréstimos, o que pode levar a uma redução do consumo e do investimento.

O desenvolvimento deste estudo contou com a vinda da pandemia e junto com ela o aumento dos preços, o trabalho se utilizou apenas de MySQL, excel, GitHub para versionamento e dados vindo dos próprios sites de órgão como o BACEN e o DIEESE

Na etapa de preparação de dados foram aplicadas algumas técnicas de limpeza e ajuste dos dados, as quais foram fundamentais para o cenário em análise, pois o mesmo possui particularidades quanto ao nome dos meses. Esses foram padronizados e mantidos como por exemplo as palavras: jan, fev etc.

As análises deste trabalho foram realizadas sobre todas as propostas cadastradas no BACEN e DIEESE entre 2020 e 2023. É importante lembrar que os valores da cesta básica estão em 'reais' por mês e a Selic em base 256 de forma mensal, com valor realizado da cesta básica do DF e Selic referente ao Brasil inteiro.

No capítulo 1 foram levantadas duas hipóteses. A primeira afirmava sobre a relação com a Selic e o papel crucial na determinação das taxas de juros praticadas pelas instituições financeiras em suas operações de crédito.

E perceber o cenário onde as pessoas e empresas podem se sentir desencorajadas a fazer empréstimos, o que pode levar a uma redução do

5. Referências

- [1]. <https://www.cnnbrasil.com.br/economia/taxa-selic-o-que-e-como-e-definida-e-quais-fatores-influenciam-se-u-sobe-e-desce/#:~:text=Quando%20a%20taxa%20Selic%20est%C3%A1,do%20consumo%20e%20do%20investimento.>
- [2]. <https://github.com/edilms/LaboratorioInovacao/tree/main>

APÊNDICE A

Arquivos de dados

base_taxa_Selic	
data	Campo que contém as datas de mês a mês formato MM/AAAA
valor	Campo que contém os valores da taxa Selic ao mês na base 252

taxa selic.csv

base_cesta_basica	
data	Campo que contém as datas de mês a mês formato MM/AAAA
valor	Campo que contém os valores da cesta básica ao mês em reais

cesta basica.csv