**Projeto de Pesquisa e Planejamento de Atividades**

|  |  |  |
| --- | --- | --- |
| **Aluno**: Gustavo de Paula Aguiar Araújo | | **Data início curso**:  25/04/2023 |
| **Orientador**: Diego De Oliveira Da Cunha | | **Defesa em:**  10/2024 |
| **Curso**: MBA Data Science e Analytics | **Modalidade**: Distância | Turma: 231 |

1. **Título do projeto**

**Desafios na identificação automática de grupos econômicos para aprimorar a gestão de riscos e controle de crédito.**

1. **Introdução**

Há muito tempo, instituições financeiras [IF]s e não financeiras têm se preocupado com a gestão eficaz do risco de crédito. Os eventos complexos que resultaram em impactos significativos nas economias e sociedades, tais como, recessões e crises financeiras, imprimiram desafios que forçaram a necessidade de controlar e desenvolver estratégias para garantir uma menor exposição ao risco de crédito, maximização dos lucros e manter a sustentabilidade no mercado. (Van Gestel e Baesens, 2008)

Uma das principais atividades bancárias é a concessão de crédito, a qual possibilita a expansão natural do mercado através da troca de recursos financeiros entre provedor e tomador. O fato da exposição ao risco de inadimplência, ou seja, do descumprimento do acordo de retorno dos recursos financeiros emprestados ao tomador pelo provedor, exige ao provedor conhecer melhor o tomador, antes de assumir o risco da transferência do recurso financeiro. Dessa forma, a necessidade de controlar e garantir o menor risco de conceção de crédito é fundamental e essencial para garantir a saúde do ambiente financeiro atual e futuro. (Van Gestel e Baesens, 2008)

A gestão de risco de crédito, ao longo do tempo, teve diferentes marcos que contribuíram para seu amadurecimento. Até o início do século XX, a análise e aprovação de crédito ainda era feita subjetivamente, dependendo somente do julgamento de analistas. Esse método, além de não utilizar critérios objetivos, era moroso e não considerava uma análise ampla, com todas as variáveis, da exposição ao risco de crédito para as IFs, tornando-se de certa forma subjetivo (Camargos, 2012). Desde então, através do advento de novas tecnologias e modelos estatísticos, as IFs vêm desenvolvendo maneiras mais eficientes para controlar sua carteira de crédito, criando mecanismos para conhecer melhor seus clientes.

É nesse contexto em que é introduzido o conceito de grupo econômico [GE]. GE é o termo dado ao conjunto de empresas que, estão interligadas por relações contratuais, cuja propriedade pertence a indivíduos ou instituições, que exercem o controle efetivo sobre essas empresas (Gonçalves, 1991). A identificação e correlação dos indivíduos de um GE é de grande importância aprimorar os modelos de precificação de risco de crédito, garantindo uma maior pluralização nas informações dos envolvidos na concessão de crédito.

Desta forma, a pergunta que orientará a pesquisa desse trabalho é: ‘Quais são as dificuldades e desafios enfrentados pelas empresas para conseguirem identificar as relações entre seus clientes e empresas, para garantir melhor condução na avaliação de risco de crédito?’.

1. **Objetivo**

Dados sempre foi de suma importância para tomar qualquer decisão, tanto pessoal como para as empresas. Desta forma, a identificação de GEs com base na participação societária é essencial no contexto de modelos de concessão de crédito e precificação de riscos. Hoje, cabe ao provedor conhecer e precificar o risco ao disponibilizar recursos financeiros ao tomador.

A falta de ferramentas eficazes e automatizadas para realizar essa identificação pode levar a lacunas na compreensão dos relacionamentos entre empresas e seus sócios, dificultado a identificação de correlação entre indivíduos que possuem algum tipo de controle efetivo em uma única empresa.

Desta forma, o propósito deste trabalho é avaliar as dificuldades para conseguir desenvolver e propor um algoritmo automatizado capaz de identificar os relacionamentos entre empresas e seus sócios.

Para atingir o propósito do projeto, este trabalho tem os seguintes objetivos específicos:

1. Identificar fontes de informações de participação societárias de empresas.
2. Avaliar as regulações propostas pelos órgãos responsáveis e sua a incorporação aos modelos de identificação de GEs.
3. Apresentar como a identificação de relações entre sócios e empresas, pode trazer maior profundidade na avaliação de risco de crédito para as IFs.
4. Demonstrar, por meio de um algoritmo, as dificuldades e necessidades para estruturação de GEs.
5. **Material e Métodos**

O projeto organiza e criar um algoritmo para conseguir disponibilizar uma visão, até certo nível, da árvore de participação societária de um GE.

Os dados privados de cada indivíduo serão preservados e somente serão utilizados os dados públicos de participação societária de empresas disponibilizados pelo Ministério da Economia [ME], os quais estão disponibilizados em diversas fontes, como os sites basedosdados.org ou dados.gov.br.

Após a identificação das bases, serão aplicadas técnicas de Data Wrangling, etapa fundamental para organizar os dados que serão utilizados para a modelagem do algoritmo, utilizando a linguagem Python ou R, ambas com poder de processamento e modelagem de datasets que nada mais são representações de dados residentes na memória os quais fornecem um modelo de programação relacional consistente, independentemente da fonte de dados (Microsoft, 2023).

Por fim, será criado um algoritmo para identificação das ligações entre sócios e empresas, utilizando grafos e, se necessário, buscas em árvores para conseguir definir a intersecção entre sócios e empresas coligadas que formam um GE.

1. **Resultados Esperados**

Espera-se que, ao final deste trabalho, seja possível identificar as dificuldades na elaboração do algoritmo automatizado para a formação de GEs. Além disso, busca-se demonstrar as etapas e processos envolvidos em sua construção, assim como a dificuldade na identificação das principais fontes de informações sobre participação societária de empresas. Pretende-se também identificar as regulações propostas pelos órgãos responsáveis e avaliar a incorporação aos modelos de identificação de grupos econômicos (GEs).

Em relação à identificação de relações entre sócios e empresas, espera-se demonstrar como essa identificação pode trazer maior profundidade na avaliação de risco de crédito para as Instituições Financeiras (IFs). Isso inclui apresentar conceitos básicos dos modelos de risco de crédito e como a identificação de GEs pode trazer melhora a avaliação das IFs.

1. **Cronograma de Atividades**

*(Adicionar as “Atividades planejadas”, assim como o período (tempo para desenvolver cada atividade) planejado para a realização de cada atividade, sendo que deverá ser adequado ao calendário de entregas das etapas do trabalho de conclusão de curso definido pela Coordenação. Marcar com um “x” a coluna que corresponde ao período planejado para desenvolver cada atividade planeja)*

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Atividades planejadas** | **Mês** | | | | | | | | | |
| **Fev** | **Mar** | **Abr** | **Mai** | **Jun** | **Jul** | **Ago** | **Set** | **Out** | **Nov** |
| Definição do tema | **x** |  |  |  |  |  |  |  |  |  |
| Envio do projeto de pesquisa |  | **x** |  |  |  |  |  |  |  |  |
| Feedbacks, correções - envio |  | **x** |  |  |  |  |  |  |  |  |
| Levantamento de bases e bibliografia |  |  | **x** |  |  |  |  |  |  |  |
| Modelar bases e tratamento de dados |  |  | **x** | **x** |  |  |  |  |  |  |
| Criar modelo de criação de grupos |  |  |  | **x** | **x** |  |  |  |  |  |
| Avaliar dificuldades e aprimorar modelo |  |  |  |  | **x** | **x** |  |  |  |  |
| Envio do resultado prévio |  |  |  |  | **x** | **x** |  |  |  |  |
| Redigir trabalho |  |  |  | **x** | **x** | **x** |  |  |  |  |
| Prévia do trabalho finalizado |  |  |  |  |  | **x** |  |  |  |  |
| Retorno com feedback |  |  |  |  |  | **x** | **x** |  |  |  |
| Finalização e entrega do trabalho escrito |  |  |  |  |  |  |  | **x** | **x** |  |
| Montar apresentação de conclusão |  |  |  |  |  |  |  |  | **x** | **x** |
| Banca – avaliação |  |  |  |  |  |  |  |  |  | **x** |
| Revisão pós banca |  |  |  |  |  |  |  |  |  | **x** |

Projeto de Pesquisa; Resultados Preliminares; Entrega do Trabalho de Conclusão de Curso; Entrega da Apresentação da Defesa

1. **Referências Bibliográficas**

Basedosdados.org. Quadros societários e de administradores das pessoas jurídicas brasileiras. Disponível em: <https://basedosdados.org/dataset/e43f0d5b-43cf-4bfb-8d90-c38a4e0d7c4f?table=81272674-f522-4e43-a70b-05bf46f0a163>. Acessado em: 24 de março de 2024.

Camargos, M.A.; Araújo, E.A.T.; Camargos, M.C.S. 2012. A inadimplência em um programa de crédito de uma instituição financeira pública de minas gerais: uma análise utilizando regressão logística. REGE - Revista de Gestão, 19(3): 473-492. Disponível em: <https://www.sciencedirect.com/science/article/pii/S1809227616303204> . Acessado em: 21 fev. 2021

Dados.gov. Cadastro Nacional da Pessoa Jurídica – CNPJ. Disponível em: <https://dados.gov.br/dados/conjuntos-dados/cadastro-nacional-da-pessoa-juridica---cnpj>. Acessado em: 24 de março de 2024.

Gonçalves, Reinaldo. 1991. Grupos econômicos: uma análise conceitual e teórica. Disponível em: <https://periodicos.fgv.br/rbe/article/view/534)>. Acessado em: 24 de março de 2024.

IME. Grafos. Disponível em: <https://www.ime.usp.br/~pf/algoritmos_para_grafos/aulas/graphs.html>.

Acessado: em 24 de março de 2024.

Microsoft. DataSets ADO.NET. Disponível em: <https://learn.microsoft.com/pt-br/dotnet/framework/data/adonet/ado-net-datasets>. Acessado em: 24 de março de 2024.

Van Gestel, T; Baesens, B. 2008. Credit Risk Management: Basic Concepts financial risk componentes, rating analysis, models, economic and regulatory capital.

Disponível em: <https://www.academia.edu/37069057/Credit_Risk_Management_Basic_Concepts>

Acessado em: 01 de abril de 2024.

<https://dados.gov.br/dados/conjuntos-dados/cadastro-nacional-da-pessoa-juridica---cnpj>  
Interface gráfica do usuário, Texto, Aplicativo, Email

Descrição gerada automaticamente

https://dadosabertos.rfb.gov.br/CNPJ/Interface gráfica do usuário, Aplicativo

Descrição gerada automaticamente

Interface gráfica do usuário, Texto, Aplicativo

Descrição gerada automaticamente

SELECT max(data) FROM `basedosdados.br\_me\_cnpj.empresas` --2023-11-16

select max(data) from `basedosdados.br\_me\_cnpj.socios` --2023-11-16

select max(data) from `basedosdados.br\_me\_cnpj.estabelecimentos` --2023-11-16

select

(select count(1) FROM `basedosdados.br\_me\_cnpj.empresas` where data = '2023-11-16') as a,

(select count(1) FROM `basedosdados.br\_me\_cnpj.socios` where data = '2023-11-16') as b

basedosdados.br\_me\_cnpj.empresas = 55.580.757 = A

basedosdados.br\_me\_cnpj.socios = 23.453.526 = B

SELECT \* FROM `basedosdados.br\_me\_cnpj.dicionario`

where id\_tabela in ('socios', 'empresas')

limit 10000

empresas = qualificacao\_responsavel = 49

empresas = porte = 5

socios = qualificacao = 22

SELECT \* FROM `basedosdados.br\_me\_cnpj.empresas`

where data = (SELECT max(data) FROM `basedosdados.br\_me\_cnpj.empresas`)

and cnpj\_basico = '57173437'

LIMIT 1000

SELECT \* FROM `basedosdados.br\_me\_cnpj.socios`

where data = (SELECT max(data) FROM `basedosdados.br\_me\_cnpj.empresas`)

and cnpj\_basico = '57173437'

LIMIT 1000

SELECT a.\*, b.\*

FROM `basedosdados.br\_me\_cnpj.empresas` as a

left join `basedosdados.br\_me\_cnpj.socios` as b on a.cnpj\_basico = b.cnpj\_basico

where a.data = “2023-11-16”

and b.data = “2023-11-16”

and a.cnpj\_basico = '57173437'

SELECT \* FROM `basedosdados.br\_me\_cnpj.estabelecimentos`

where data = (SELECT max(data) FROM `basedosdados.br\_me\_cnpj.empresas`)

LIMIT 1000

SELECT \* FROM `basedosdados.br\_me\_cnpj.simples`

where cnpj\_basico = '30032502'

LIMIT 1000

Select para avaliar se a cota excedeu

SELECT

 job\_id,

 creation\_time,

 error\_result

FROM `region-us`.INFORMATION\_SCHEMA.JOBS

WHERE creation\_time > TIMESTAMP\_SUB(CURRENT\_TIMESTAMP, INTERVAL 1 DAY) AND

      error\_result.reason IN ('rateLimitExceeded', 'quotaExceeded')

casos que possuem mais de uma relação dentro do banco – possíveis grafos integrados  
select count(1), d.documento, d.nome

from (

  SELECT distinct c.documento,c.nome, a.cnpj\_basico

  FROM basedosdados.br\_me\_cnpj.empresas a

  inner join basedosdados.br\_me\_cnpj.estabelecimentos b on a.cnpj\_basico = b.cnpj\_basico and b.situacao\_cadastral in ("1","2","4") and b.identificador\_matriz\_filial = "1"

  inner join basedosdados.br\_me\_cnpj.socios c on a.cnpj\_basico = c.cnpj\_basico and c.data = "2023-11-16"

  where a.data = "2023-11-16" and b.data = "2023-11-16"

  limit 1000000

) as d

group by  d.documento, d.nome

having count(1) > 3

BIGQUERY – google  
<https://console.cloud.google.com/bigquery?p=basedosdados&d=br_me_cnpj&t=dicionario&page=table&project=gruposeconomicos&ws=!1m14!1m4!4m3!1sbasedosdados!2sbr_me_cnpj!3sdicionario!1m8!1m3!1sgruposeconomicos!2sbquxjob_657a905f_18fb2d54bef!3sUS!14m3!1sgruposeconomicos!2sbquxjob_4aad6b92_18fb2d80445!3sUS>  
  
Pra acessar, avaliar o vídeo <https://www.youtube.com/watch?v=z32438Yehl4>  
e depois de logado, clicar no botão azul da base.   
Interface gráfica do usuário, Texto, Email

Descrição gerada automaticamente

Valida situação cadastral da empresa.   
[https://solucoes.receita.fazenda.gov.br/Servicos/cnpjreva/Cnpjreva\_Solicitacao.asp?cnpj=59275289000102Interface gráfica do usuário, Texto, Aplicativo, Word

Descrição gerada automaticamenteInterface gráfica do usuário, Texto, Aplicativo

Descrição gerada automaticamente](https://solucoes.receita.fazenda.gov.br/Servicos/cnpjreva/Cnpjreva_Solicitacao.asp?cnpj=59275289000102)

Interface gráfica do usuário, Texto, Aplicativo

Descrição gerada automaticamente

Python

<https://medium.com/basedosdados/intro-ao-pacote-basedosdados-em-python-4e05439e936d>

Interface gráfica do usuário, Texto, Aplicativo

Descrição gerada automaticamente

<https://github.com/basedosdados/mais?tab=readme-ov-file#usando-em-python>

Configurando conta de serviço BigQuery:  
https://www.youtube.com/watch?v=vz4dg9lxB1Y  
Interface gráfica do usuário, Texto, Aplicativo, Email

Descrição gerada automaticamente

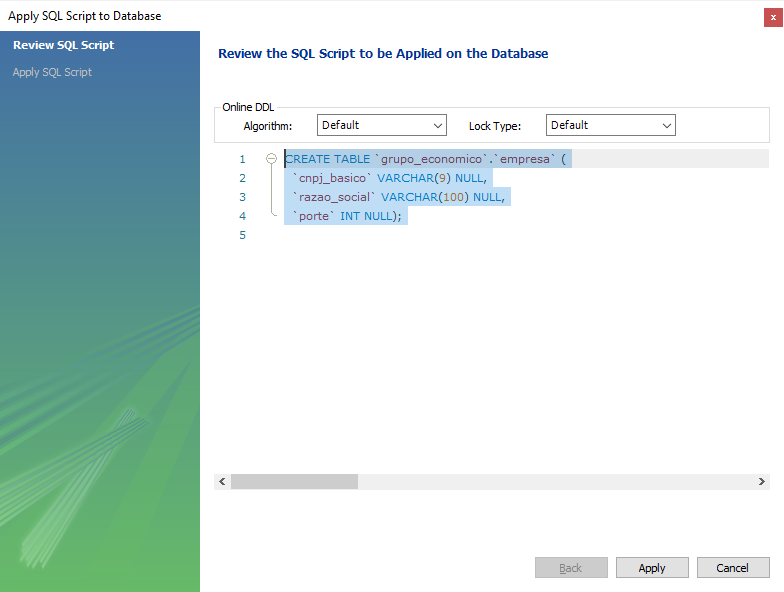
Interface gráfica do usuário, Texto, Aplicativo, Email

Descrição gerada automaticamente

MYSQL  
CREATE TABLE `grupo\_economico`.`empresas` (

`cnpj\_basico` VARCHAR(9) NULL,

`razao\_social` VARCHAR(100) NULL,

`porte` INT NULL);  


CREATE TABLE `grupo\_economico`.`socios` (

`cnpj\_basico` VARCHAR(9) NULL,

`tipo` INT NULL,

`nome` VARCHAR(100) NULL,

`documento` VARCHAR(14) NULL,

`qualificacao` INT NULL,

`faixa\_etaria` INT NULL);  
Interface gráfica do usuário, Aplicativo

Descrição gerada automaticamente  
  
  
CREATE TABLE `grupo\_economico`.`estabelecimentos` (

`cnpj\_basico` VARCHAR(9) NULL,

`cnpj` VARCHAR(14) NULL,

`nome\_fantasia` VARCHAR(100) NULL,

`identificador\_matriz\_filial` INT NULL,

`situacao\_cadastral` INT NULL);  
Interface gráfica do usuário, Aplicativo

Descrição gerada automaticamente

CREATE TABLE `grupo\_economico`.`dicionario` (

`id\_tabela` VARCHAR(50) NOT NULL,

`nome\_coluna` VARCHAR(100) NOT NULL,

`chave` INT NOT NULL,

`valor` VARCHAR(100) NULL); Interface gráfica do usuário, Aplicativo

Descrição gerada automaticamente

Texto

Descrição gerada automaticamente  
  
  
  
Interface gráfica do usuário, Aplicativo

Descrição gerada automaticamente CREATE TABLE `grupo\_economico`.`empresa\_socios` (

`cnpj\_basico` VARCHAR(9) NULL,

`razao\_social` VARCHAR(200) NULL,

`porte` INT NULL,

`tipo\_socio` INT NULL,

`nome\_socio` VARCHAR(200) NULL,

`doc\_socio` VARCHAR(14) NULL,

`qualificacao\_socio` INT NULL,

`faixa\_etaria\_socio` INT NULL,

INDEX `cnpj\_base` (`cnpj\_basico` ASC) INVISIBLE,

INDEX `nome\_doc` (`doc\_socio` ASC, `nome\_socio` ASC) VISIBLE);

Interface gráfica do usuário, Texto, Aplicativo

Descrição gerada automaticamente

CREATE TABLE `grupo\_economico`.` empresa\_socios\_grafo` (

`empresas` VARCHAR(250) NULL,

`socios` VARCHAR(250) NULL,

`id\_grupo` INT NULL,

INDEX `empresa` (`empresas` ASC) VISIBLE,

INDEX `socio` (`socios` ASC) VISIBLE,

INDEX `id` (`id\_grupo` ASC) VISIBLE,

INDEX `empresa\_socio` (`empresas` ASC, `socios` ASC) VISIBLE);

Interface gráfica do usuário, Texto, Aplicativo

Descrição gerada automaticamente

CREATE TABLE `grupo\_economico`.`empresa\_socios\_pre\_grafo` (

`empresas` VARCHAR(250) NULL,

`socios` VARCHAR(250) NULL,

INDEX `empresas` (`empresas` ASC) INVISIBLE,

INDEX `socios` (`socios` ASC) INVISIBLE,

INDEX `emp\_soc` (`empresas` ASC, `socios` ASC) VISIBLE);

import data  
Tela de computador

Descrição gerada automaticamente