**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.

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.

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. (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 este conjunto de empresas (Gonçalves, 1991). A identificação e correlação dos indivíduos de um GE são de grande importância, para conseguir introduzir mais informações aos modelos de precificação de risco de crédito, garantindo uma pluralização maior nas informações dos envolvidos na concessão 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 de suma importância 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.

Atualmente, 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 desenvolver um algoritmo automatizado capaz de identificar os relacionamentos entre empresas e seus sócios, utilizando a participação societária entre eles, demonstrando que existe uma correlação na identificação na formação de GEs e modelos de precificação de risco mais assertivos.

1. **Material e Métodos**

A ideia principal do projeto é organizar e criar um algoritmo que consiga disponibilizar uma visão, até certo nível, da árvore de participação societária de um grupo econômico. Desta forma, o desafio inicial é conseguir obter uma massa de dados para provar o modelo.

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 o site de 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**

*(Descrever os resultados que são esperados após a realização da coleta e análise dos dados, ou seja, quais resultados são esperados ao final da pesquisa)*

Espera-se ao final deste trabalho, que seja possível identificar uma maneira eficiente de organizar os dados de participação societárias das empresas para conseguir identificar de forma eficiente a formação dos GEs.

Também será demonstrado e criado o algoritmo para formação dos grupos, os quais poderão ser utilizados como insumos para a modelagem de modelos de risco de créditos mais apurados, visto que a ligação entre indivíduos traz um enriquecimento das características dos envolvidos nas transações bancárias.

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.