



**UNIVERSIDADE FEDERAL DO ACRE**  
**CURSO DE BACHARELADO EM SISTEMAS DE INFORMAÇÃO**

**ANÁLISE DO IMPACTO DA LOCALIZAÇÃO GEOGRÁFICA DOS  
COLABORADORES NA ACEITAÇÃO E TEMPO DE VIDA DE *PULL REQUESTS***

**RIO BRANCO**  
**2019**

**CLEYCIANE FARIAS DE LIMA**

**ANÁLISE DO IMPACTO DA LOCALIZAÇÃO GEOGRÁFICA DOS  
COLABORADORES NA ACEITAÇÃO E TEMPO DE VIDA DE *PULL REQUESTS***

Monografia apresentada como exigência final para obtenção do grau de bacharel em Sistemas de Informação da Universidade Federal do Acre.

Prof. Orientador: Dr. Daricélio Moreira Soares

**RIO BRANCO**

**2019**

## **TERMO DE APROVAÇÃO**

**CLEYCIANE FARIAS DE LIMA**

### **ANÁLISE DO IMPACTO DA LOCALIZAÇÃO GEOGRÁFICA DOS COLABORADORES NA ACEITAÇÃO E TEMPO DE VIDA DE *PULL REQUESTS***

Esta monografia foi apresentada como trabalho de conclusão de Curso de Bacharelado em Sistemas de Informação da Universidade Federal do Acre, sendo aprovado pela banca constituída pelo professor orientador e membros abaixo mencionados.

Compuseram a banca:

---

Prof. Dr. Daricélio Moreira Soares (Orientador)  
Curso de Bacharelado em Sistemas de Informação

---

Prof(a) Dra. Catarina de Souza Costa  
Curso de Bacharelado em Sistemas de Informação

---

Prof. Dr. Olacir Rodrigues Castro Júnior  
Curso de Bacharelado em Sistemas de Informação

Rio Branco - AC, 12 de Dezembro de 2019

*Dedico este trabalho a Deus, minha família e Amigos.*

## **AGRADECIMENTOS**

No decorrer da minha vida acadêmica foram enfrentadas muitas dificuldades, que foram vivenciadas no dia a dia, nas quais tenho a honra em dizer que superei cada uma delas com a ajuda de Deus e da minha família, por mais difíceis que tenham sido.

Primeiramente quero agradecer a Deus pelo privilégio de estar escrevendo esta seção de agradecimentos, sem Ele não teria chegado até aqui, e obtido o primeiro e o mais importante passo da minha vida, a graduação.

Agradeço a minha família, em especial ao meu pai, Valdir Farias Vieira Filho, minha mãe, Eliene de Lima Moura e minha irmã Glenna Farias de Lima, pelo apoio, paciência e carinho que tiveram, pela preocupação, por sempre acreditarem em mim, e me incentivarem a sempre seguir em frente e nunca desistir, pelos ensinamentos e educação. Vocês serão sempre minha base, amo vocês e agradeço a Deus pela vida de cada um.

Ao meu orientador, professor Drº Daricélio Soares, pela paciência e dedicação que teve ao me orientar e tirar minhas dúvidas, serei eternamente grata ao senhor, pelos seus ensinamentos que levarei por toda a vida.

Aos meus professores que tive durante toda a graduação, meus eternos agradecimentos, pelo apoio e dedicação que sempre demonstraram ao fazerem o que amam, que é lecionar.

Aos meus colegas e amigos de graduação que tive a honra de estudar com eles, Gabriel Figueiredo, Mateus Costa, Salomão Mafalda e em especial minha amiga Juliana Abreu da Cunha, muito obrigada pelos ensinamentos, compreensão e paciência que você teve, por aceitar muitas das vezes meus desafios e acreditar na realização deles, pelas viagens, pelas noites de sono perdidas muitas das vezes para terminar projetos, ou estudar para as provas, meus eternos agradecimentos.

Ao time das maratonas de programação que participei durante a graduação e ao professor/coach dos times, Dr. Manoel Limeira, obrigada pelo incentivo, pelos momentos de aprendizado que levarei por toda a vida. Não poderia deixar de agradecer as minhas amigas Beatriz Gomes, Juliana Abreu, Tainá Cavalcante, por aceitarem fazer parte do time; IRMÃO OLHO, ٩( ^。 ^ ) J whatever, Tio Limeira ++, Pode Ser, meus eternos agradecimentos pois como equipe garantimos algumas classificações regionais e uma Final Brasileira da maratona de programação em 2017 em Foz do Iguaçu. Não poderia ter feito parte de outra equipe que não fossem vocês, além de equipe nos tornamos amigas, algo que levarei para a vida.

Aos meus amigos que me apoiaram nesta etapa e oraram por mim, meus eternos agradecimentos, pois cada um teve uma parte importante em minha vida e isso serei eternamente grata.

A Silvana de Andrade Gonçalves, por tirar minhas dúvidas, ler meus trabalhos e sempre me ajudar principalmente na etapa final deste trabalho, meus eternos agradecimentos.

Agradeço pela Universidade Federal do Acre por ter me dado a oportunidade de estudar e concluir este curso.

Agradeço a todos os que fizeram parte direta ou indiretamente na minha vida, para que a minha conclusão do curso fosse possível.

*“Porque sem mim nada podeis fazer. ”*

*João 15:5*

## RESUMO

O Desenvolvimento Distribuído de Software - DDS, permite trabalhar com desenvolvimento de software de forma global unindo as pessoas de diferentes localidades para a realização de software em conjunto. Uma das principais características do DSS consiste na utilização de contribuições externas que podem ser sistematizadas através de um paradigma de colaboração chamado de *Pull request*. Nesse paradigma colaboradores externos criam um *fork* a partir do repositório, fazem sua integração e enviam um *Pull Request* o qual a equipe decidirá se irá integrar ou não ao repositório principal. Atualmente a literatura tem buscado identificar quais são os fatores que influenciam na aceitação e rejeição de *Pull Requests*, sendo assim, este trabalho tem por objetivo identificar como a localização geográfica influencia na aceitação e tempo de vida do *Pull Request* e o relacionamento com fatores físicos das contribuições.

Palavras-chave: *Pull Request*, Desenvolvimento Distribuído de Software, Localização Geográfica.



## ABSTRACT

The Distributed Software Development (DDS) allows the work with Software Development in a global way, by uniting people from different locations of the world to make software together. One of the main features of DSS is the use of external contributions that can be systematized through a collaboration paradigm called Pull request. In this Paradigm, external contributors make a fork from the repository, integrate it and send a Pull Request in which the team will decide, whether or not, to integrate into the main repository. Currently, the literature has sought to identify the factors that influence the acceptance and rejection of *Pull Requests*. Therefore, this work aims to identify how the geographical location influences the acceptance and *lifetime* of Pull Request and the relationship with physical factors of *Pull Requests* contributions.

Key-words: *Pull Request*, Distributed Development Software, Geographic location.

## LISTAS DE FIGURAS

Figura 1 - Metodologia .....	22
Figura 2 - Mesma Localização Física .....	26
Figura 3 - Distância Nacional .....	26
Figura 4 - Distância Continental .....	27
Figura 5 - Distância Global .....	27
Figura 6 - Processo de Gerencia de Configuração .....	28
Figura 7 - Modelo Centralizado .....	31
Figura 8 - Modelo Distribuído .....	32
Figura 9 - Processo de Pull .....	33
Figura 10 - Tela de configuração do algoritmo Apriori .....	43
Figura 11 – <i>Lift</i> da regra do tipo: <i>Country</i> → <i>Status</i> = "closed" .....	44
Figura 12 – <i>Lift</i> da regra do tipo: <i>Country</i> → <i>Status</i> = "merged" .....	45
Figura 13 – <i>Lift</i> da regra do tipo: <i>Country</i> → <i>lifetimeMinutes</i> = "lengthy" .....	46
Figura 14 – <i>Lift</i> da regra do tipo: <i>Country</i> → <i>lifetimeMinutes</i> = "medium" .....	47
Figura 15 – <i>Lift</i> da regra do tipo: <i>Country</i> → <i>lifetimeMinutes</i> = "short" .....	48
Figura 16 – <i>Lift</i> da regra do tipo: <i>Country</i> → <i>lifetimeMinutes</i> = "very lengthy" .....	48
Figura 17 – <i>Lift</i> da regra do tipo: <i>Country</i> → <i>lifetimeMinutes</i> = "very short" .....	49

## LISTAS DE QUADROS

Quadro 1 - Atributos para a extração das regras.....	41
--	----

## LISTAS DE TABELAS

Tabela 1 - Base de dados com 03 transações .....	37
Tabela 2 - Cálculo do Suporte.....	38
Tabela 3 - Cálculo da confiança e <i>lift</i> da regra .....	38
Tabela 4 – Países com maiores probabilidades de aceitação quando <i>Country</i> → <i>changedFiles</i> = “1 File”.....	51
Tabela 5 - Países com menores probabilidades de aceitação quando <i>Country</i> → <i>changedFiles</i> = “1 File”.....	51
Tabela 6 - Países com maiores probabilidades de aceitação quando <i>Country</i> → <i>changedFiles</i> = “Many Files” .....	52
Tabela 7 - Países com menores probabilidades de aceitação quando <i>Country</i> → <i>changedFiles</i> = “Many Files”. .....	53
Tabela 8 - Países com maiores probabilidades de aceitação quando <i>Country</i> → <i>changedFiles</i> = “Some Files”.....	53
Tabela 9 - Países com menores probabilidades de aceitação quando <i>Country</i> → <i>changedFiles</i> = “Some Files”.....	54
Tabela 10 - Países com maiores probabilidades de aceitação quando <i>Country</i> → <i>totalLines</i> = “Many Lines” .....	54
Tabela 11 - Países que possuem menores chances de aceitação quando <i>Country</i> → <i>totalLines</i> = “Many Lines” .....	55
Tabela 12 - Países que possuem maiores chances de aceitação quando <i>Country</i> → <i>totalLines</i> = “Some Lines”.....	55

Tabela 13 - Países que possuem menores chances de aceitação quando <i>Country</i> → <i>totalLines</i> = " <i>Some Lines</i> ".....	56
Tabela 14 - Regras de associação relevantes para aceitação quando houverem modificações de arquivos em Singapura.....	57
Tabela 15 - Regras de associação relevantes para aceitação quando houverem modificações de arquivos em Belarus.....	58
Tabela 16 - Regras de associação relevantes para rejeição quando houverem modificações de arquivos em Itália .....	58
Tabela 17 - Regras de associação relevantes para rejeição quando houverem modificações de arquivos em Reino Unido .....	59
Tabela 18 - Regras de associação relevantes para aceitação quando houverem modificações de arquivos em República Checa .....	59
Tabela 19 – Regras de associação relevantes para aceitação quando houverem modificações de linhas em Singapura. ....	60
Tabela 20 - Regras de associação relevantes para aceitação quando houverem modificações de linhas em Belarus. ....	60
Tabela 21 – Regras de associação relevantes para aceitação quando houverem modificações de linhas na Itália.....	61
Tabela 22 - Regras de associação relevantes para aceitação quando houverem modificações de linhas no Reino Unido.....	61
Tabela 23 - Regras de associação relevantes para aceitação quando houverem modificações de linhas na República Checa .....	62
Tabela 24 - Regras de associação relevantes para <i>lifetime</i> quando houverem modificações de 1 arquivo em Singapura.....	62
Tabela 25 - Regras de associação relevantes para <i>lifetime</i> quando houverem modificações de 1 arquivo em Belarus.....	63
Tabela 26 - Regras de associação relevantes para <i>lifetime</i> quando houverem modificações de 1 arquivo na Itália .....	64
Tabela 27 - Regras de associação relevantes para <i>lifetime</i> quando houverem modificações de muitos arquivos na Itália .....	64
Tabela 28 - Regras de associação relevantes para <i>lifetime</i> quando houverem modificações de alguns arquivos na Itália. ....	65

Tabela 29 - Regras de associação relevantes para <i>lifetime</i> quando houverem modificações de arquivos na República Checa. ....	65
Tabela 30 - Regras de associação relevantes para <i>lifetime</i> quando houverem modificações de arquivos no Reino Unido. ....	66
Tabela 31 - Regras de associação relevantes para <i>lifetime</i> quando houverem modificações de linhas em Singapura. ....	67
Tabela 32 - Regras de associação relevantes para <i>lifetime</i> quando houverem modificações de linhas em Belarus. ....	67
Tabela 33 - Regras de associação relevantes para <i>lifetime</i> quando houverem modificações de linhas em países com índices de rejeição. ....	68
Tabela 34 - <i>Country ^ Language</i> → <i>Status</i> = " <i>Merged</i> " .....	69
Tabela 35 - <i>Country ^ Language</i> → <i>Status</i> = " <i>Closed</i> " .....	69

## SUMÁRIO

<b>1 INTRODUÇÃO.....</b>	<b>17</b>
<b>1.1 PROBLEMA DA PESQUISA .....</b>	<b>18</b>
<b>1.2 OBJETIVOS DA PESQUISA .....</b>	<b>19</b>
1.2.1 Objetivo Geral .....	20
1.2.2 Objetivos Específicos .....	20
<b>1.3 JUSTIFICATIVA DA PESQUISA .....</b>	<b>20</b>
<b>1.4 METODOLOGIA .....</b>	<b>21</b>
<b>1.5 ORGANIZAÇÃO DO ESTUDO .....</b>	<b>22</b>
<b>2 FUNDAMENTAÇÃO TEÓRICA.....</b>	<b>24</b>
<b>2.1 DESENVOLVIMENTO DISTRIBUÍDO DE SOFTWARE - DDS.....</b>	<b>25</b>
<b>2.2 GERÊNCIA DE CONFIGURAÇÃO - GC.....</b>	<b>28</b>
<b>2.3 SISTEMAS DE CONTROLE DE VERSÃO - SCV .....</b>	<b>30</b>
2.3.1 MODELO CENTRALIZADO E DISTRIBUÍDO .....	31
<b>2.4 <i>PULL REQUESTS</i>.....</b>	<b>33</b>
<b>2.5 ACEITAÇÃO DE <i>PULL REQUESTS</i> EM PROJETOS OPEN SOURCE....</b>	<b>34</b>
<b>2.6 REGRAS DE ASSOCIAÇÃO .....</b>	<b>35</b>
<b>2.7 CONSIDERAÇÕES FINAIS SOBRE O CAPÍTULO.....</b>	<b>39</b>
<b>3 RESULTADOS E DISCUSSÕES.....</b>	<b>40</b>
<b>3.1 INFLUÊNCIA DA LOCALIZAÇÃO GEOGRÁFICA DO <i>REQUESTER</i> NA ACEITAÇÃO <i>PULL REQUESTS</i>.....</b>	<b>40</b>
<b>3.2 CARACTERÍSTICAS DE ACEITAÇÃO DE <i>PULL REQUESTS</i>.....</b>	<b>50</b>

3.3 FATORES DE INFLUÊNCIA NO PAÍS .....	56
4 CONSIDERAÇÕES FINAIS E RECOMENDAÇÕES .....	71
4.1 CONSIDERAÇÕES FINAIS .....	71
4.2 RECOMENDAÇÕES.....	72
REFERÊNCIAS .....	73
APÊNDICE .....	76
APÊNDICE A– PROJETOS QUE COMPÕEM A BASE DE DADOS.....	77
APÊNDICE B– LOCALIZAÇÃO GEOGRÁFICA DOS CONTRIBUIDORES....	79



## 1 INTRODUÇÃO

A Gerência de Configuração de Software (GCS) é fundamental para o processo de desenvolvimento colaborativo de software, o qual, muitas vezes envolve a distribuição física da equipe de desenvolvimento, modelo conhecido como Desenvolvimento Distribuído de Software (DDS).

A partir de critérios definidos, DDS proposto por Audy e Prikladnicki (2008) aplica-se, na caracterização de um ambiente, em que pelo menos um dos atores envolvidos estejam fisicamente distantes dos demais. Nesse cenário, a contribuição que é enviada à equipe principal do projeto para avaliação é denominado *Pull Request*. De acordo com Soares (2017), colaboradores externos podem fazer modificações isoladas em artefatos e solicitar a integração de suas modificações de volta ao repositório principal do projeto.

Em projetos de código aberto, de acordo com Soares (2017), a sistematização do processo de contribuição é extremamente importante, tendo em vista que os desenvolvedores interessados no projeto podem enviar suas contribuições para os repositórios remotos destes projetos.

O envio de um *Pull Request* marca o início de sua vida útil. Durante a revisão os desenvolvedores da equipe principal podem comentar e discutir sobre um *Pull Request*, exigindo potencialmente modificações adicionais do solicitante. Essa interação pode afetar o tempo necessário para avaliar a contribuição e, como resultado, afetar o status final do *Pull Request*, ou seja, aceitação ou rejeição (SOARES, 2017).

Neste contexto, este trabalho, visa analisar o impacto da localização geográfica dos colaboradores na aceitação e tempo de vida de *Pull Request*, tendo em vista que, ao enviarem o *Pull Request* os *requesters* (colaboradores) podem estar geograficamente dispersos dos *reviewers* (revisores), podendo afetar a aceitação e rejeição do projeto.

## 1.1 PROBLEMA DA PESQUISA

Durante a evolução do software, o processo de Gerência de Configuração de Software (GCS) é fundamental, pois permite que os artefatos modificados por diferentes desenvolvedores sejam gerenciados e versionados, identificando produtos de trabalho que são susceptíveis a mudanças e estabelecendo relações entre eles. Sistemas de Controle de Versões (SCV), permitem um acompanhamento minucioso do andamento das tarefas de manutenção, possibilitando que diversas métricas sejam coletadas e analisadas (MURTA et al., 2004).

O processo de GCS surgiu para evitar a perda do controle de projeto, tendo em vista a grande quantidade de informações que são produzidas e podem ser alteradas a qualquer instante. Além disso, do ponto de vista da gerência do projeto, a GCS é importante, pois controla a evolução e a integridade de um produto, ao identificar seus elementos, gerenciar e controlar mudanças, e ao registrar e informar tais mudanças (FIGUEIREDO; ROCHA; SANTOS, 2004). De acordo com Soares (2017), em projetos de código aberto, várias pessoas podem contribuir para o desenvolvimento do software. Por exemplo, um desenvolvedor externo pode descobrir uma falha no código fonte, corrigir e em seguida solicitar a incorporação da alteração no software.

Para Kamei e Pinto (2014), a distribuição do processo de desenvolvimento de software, faz ampliar problemas inerentes ao desenvolvimento tradicional e gera novos desafios ao adicionar distâncias físicas. Algumas maneiras de enfrentar estes desafios consistem na adoção de processos, técnicas e ferramentas de suporte ao

DDS, que objetivem a efetiva comunicação e coordenação de equipes distribuídas. O principal método para contribuições de códigos consiste no *Pull Request*, onde para manter a qualidade dos projetos de software, a revisão é a parte essencial no DDS.

Ao enviarem suas modificações, os *requesters* podem estar geograficamente dispersos dos *reviewers*, podendo afetar a aceitação e rejeição do projeto, conforme indica o trabalho de Rastogi et al., (2018), onde projetos de software de código aberto, mostram viés sugerindo que outras características demográficas de desenvolvedores, como localização geográfica, podem influenciar negativamente a avaliação de contribuintes.

Considerando este contexto, este trabalho apresenta um conjunto de estudos sobre a influência da localização geográfica do *requester* nos cenários de aceitação e tempo de vida do *Pull Request* e o relacionamento com os fatores físicos das contribuições. Os estudos visam responder as seguintes questões da pesquisa:

- A localização geográfica e o *lifetime* influenciam na aceitação/rejeição de *Pull Requests*?
- Dado que um país de origem de um *Pull Request* influencia na aceitação/rejeição dos projetos, quais características estão presentes nestes *Pull Requests*?
- Quais os fatores potencializam ou enfraquecem a influência do país?

## 1.2 OBJETIVOS DA PESQUISA

Nesta seção serão descritos o objetivo geral e os objetivos específicos da pesquisa.

### 1.2.1 Objetivo Geral

Analisar a influência da localização geográfica do *requester* nos cenários de aceitação e tempo de vida do *Pull Request* e o relacionamento com fatores físicos das contribuições.

### 1.2.2 Objetivos Específicos

Para a obtenção do objetivo geral têm-se os seguintes objetivos específicos:

- a) Revisar bibliograficamente os conceitos que dão embasamento a esta pesquisa;
- b) Selecionar dados para a experimentação;
- c) Realizar pré-processamento e enriquecimento de base de dados;
- d) Extrair regras de associação sobre os fatores influentes na aceitação do tempo de vida de *Pull Requests* considerando localizações geográficas;
- e) Analisar e discutir os resultados.

## 1.3 JUSTIFICATIVA DA PESQUISA

No desenvolvimento baseado em *Pull Request*, todas as pessoas envolvidas no projeto, segundo Soares (2017), têm interesse em ter contribuições aceitas, se consideradas viáveis para o projeto.

Em geral, no cenário de colaboração, a compreensão de fatores que podem influenciar na aceitação e rejeição de *Pull Requests* são importantes, tendo em vista que de acordo com Soares (2017) a apresentação de um *Pull Request* marca o começo de sua vida útil, sendo que essa interação pode impactar o tempo necessário

para avaliar a contribuição e como resultado impactar também o *Status* final de aceitação ou rejeição do projeto.

Analisar os fatores no cenário de aceitação e tempo de vida do *Pull Request*, pode ajudar os desenvolvedores localizados em diversos lugares do mundo a orientar suas ações para aumentar as chances de aceitação, além disso o desenvolvedor pode identificar as principais características e os fatores que potencializam ou enfraquecem a influência deste país, quando submetem uma solicitação.

## 1.4 METODOLOGIA

Segundo Silva e Menezes (2005), do ponto de vista de sua natureza a pesquisa se classifica como aplicada, tendo em vista que objetiva gerar conhecimentos para aplicações práticas e dirigidos à solução de problemas específicos.

Do ponto de vista da abordagem do problema pode ser considerada como pesquisa quantitativa, e do ponto de vista de seus objetivos, segundo Gil (1991), é denominada como exploratória em que visa proporcionar maior familiaridade com os problemas com vistas a torna-la explicito ou construir hipóteses.

Do ponto de vista dos procedimentos técnicos, a pesquisa será classificada como estudo de caso, onde será envolvido o estudo profundo e exaustivo de um ou poucos objetos de maneira que se permita o seu amplo e detalhado conhecimento (GIL,2008).

Com relação aos métodos utilizados, que proporcionam as bases lógicas de investigação, essa pesquisa classifica como método dedutivo, partindo do pressuposto em que a localização geográfica tem influência na aceitação e tempo de vida de pedidos de *Pull Requests*. Quanto aos meios técnicos de investigação a pesquisa classificam-se como método comparativo, onde serão analisados arquivos com localizações dos requesters, e qual o grau de aceitação das solicitações desses requesters.

Para a realização do trabalho, foram realizadas as seguintes etapas, resumidas na Figura 1 e detalhado posteriormente:

**Figura 1 - Metodologia**



Fonte: Elaboração Própria

- a) A primeira etapa consistiu na seleção dos projetos GitHub (Projetos ativos, populares, diferentes localidades, etc.) e após isso foi feita a análise dos atributos da localização geográfica;
- b) A segunda etapa se deu na extração das regras nos projetos, e análise dos atributos da localização geográfica com relação a aceitação e tempo de vida de *Pull Request*;
- c) E por fim, a terceira consistiu na análise e discussão dos dados que foram coletados durante a pesquisa que foi realizada, para verificar se a localização influencia na aceitação e tempo de vida de um *Pull Request*.

## 1.5 ORGANIZAÇÃO DO ESTUDO

Este trabalho está organizado em 4 capítulos. Além do capítulo de introdução, no qual foram abordados o problema da pesquisa, os objetivos que fundamentam a pesquisa, justificativa e metodologia, o restante do trabalho é composto por mais três capítulos.

No Capítulo 2 são apresentados conceitos que fundamentam a pesquisa que foi realizada, sendo estes, o Desenvolvimento Distribuído de Software – DDS, a

Gerência de Configuração de Software – GCS, Sistemas de Controle de Versão – SCV abordando os principais modelos de gerenciamento de repositórios sendo estes: Modelo Centralizado e Distribuído, será abordado também o tópico de *Pull Requests* e a Aceitação de *Pull Requests* em projetos de desenvolvimento Open Source.

O Capítulo 3 apresenta o estudo de caso tendo como suporte a base de dados utilizada por Soares (2017), analisando atributos da localização geográfica e o tempo de vida, para saber se estes fatores possuem influência, quais as principais características que justifiquem a aceitação ou rejeição dos projetos e quais as regras que atenuam estes fatores.

Por fim, o Capítulo 4 apresenta as considerações finais do trabalho incluindo suas limitações e contribuições e perspectivas para trabalhos futuros.

## 2 FUNDAMENTAÇÃO TEÓRICA

Em projetos de desenvolvimento de software a aceitação ou rejeição de *Pull Requests* do repositório principal, envolvem estudos e análises para compreensão de possíveis fatores que possam influenciar no status final de aceitação desses *Pull Requests*. Sendo assim, este capítulo contém os conceitos em que fundamentam esse trabalho. Serão abordados os seguintes temas com suas respectivas sessões:

Na Seção 2.1 são apresentados os conceitos de Desenvolvimento Distribuído de Software (DDS), abordando a sua caracterização e desafios, destacando o nível de dispersão de membros de um determinado grupo de *stakeholders* (envolvidos) de um projeto de software.

A Seção 2.2 apresenta a Gerência de Configuração de Software (GCS), abordando sua importância para o controle de sistemas complexos, a fim de evitar a perda do controle do projeto de desenvolvimento de software.

Na Seção 2.3 é apresentado o tópico sobre Sistemas de Controle de Versão (SCV) descrevendo posteriormente na subseção 2.3.1 os modelos de gerenciamento de repositórios denominado como, Modelo Centralizado e Distribuído.

A Seção 2.4 descreve o paradigma de contribuição *Pull Request*, que é empregado para a sistematização de contribuições e projetos de desenvolvimento de software.



A Seção 2.5 discute a aceitação de *Pull Requests* em projetos *Open Source*, buscando compreender os fatores que podem contribuir na aceitação ou rejeição de projetos em códigos.

Por fim, a Seção 2.6 aborda Regras de associação, destacando conceito, exemplo e aplicação no contexto de desenvolvimento.

## 2.1 DESENVOLVIMENTO DISTRIBUÍDO DE SOFTWARE - DDS

O DDS de acordo com Prikladnicki et al. (2004), é caracterizado pela colaboração e cooperação entre departamentos de organizações e pela criação de grupos de pessoas, as quais estão localizadas em cidades ou países diferentes, distantes tanto temporal quanto fisicamente.

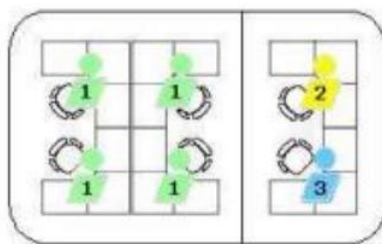
Apesar de muitas vezes o processo ocorrer em um mesmo país, em regiões com incentivos fiscais ou de concentração de massa crítica em determinadas áreas, algumas empresas buscam vantagens em soluções externas até mesmo em outros países (PRIKLADNICKI, 2003).

As razões que se destacam para a aplicação do DDS, são várias. Essas ações ou subconjuntos delas, motivam um crescente número de organizações a desenvolverem software da forma distribuída (AUDY; PRIKLADNICKI, 2007),

Entre os desafios no contexto de projetos distribuídos destacam-se o nível de dispersão de membros de um determinado grupo de *stakeholders* (envolvidos), de um projeto de software, onde Prikladnicki et al. (2004) destacam quatro situações, das quais verificam o tipo de distância física e suas características sendo elas:

**Mesma localização física:** Onde todos os *stakeholders* estão no mesmo local, as reuniões ocorrem sem dificuldade e a equipe interage estando fisicamente presente. Nesse aspecto não há diferença de fuso horário e as diferenças culturais raramente envolvem dimensão nacional (Figura 2).

**Figura 2 - Mesma Localização Física**



Fonte: Prikladnicki et al. (2004)

**Distância nacional:** Todos os *stakeholders* estão em um mesmo país, as reuniões ocorrem com certa dificuldade encontrando-se em intervalos longos com curtas durações de tempo (Figura 3).

**Figura 3 - Distância Nacional**



Fonte: Prikladnicki et al. (2004)

**Distância Continental:** Todos os *stakeholders* estão em países diferentes, porém em mesmo continente. As reuniões ocorrem com maior dificuldade encontrando-se em intervalos longos com curtas durações de tempo. Nesse aspecto podem haver diferenças de fuso horário, dificultando algumas interações como marcar reuniões (Figura 4).

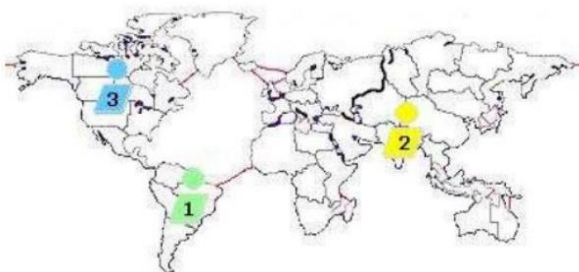
**Figura 4 - Distância Continental**



Fonte: Prikladnicki et al. (2004)

**Distância Global:** Nesse aspecto, os envolvidos no projeto (*stakeholders*) estão localizados em continentes diferentes, formando distribuições globais. Nesse caso, as reuniões raramente acontecem, tendo em vista a dificuldade de locomoção, além disso barreiras culturais e fusos horários podem ser empecilhos para o trabalho (Figura 5).

**Figura 5 - Distância Global**



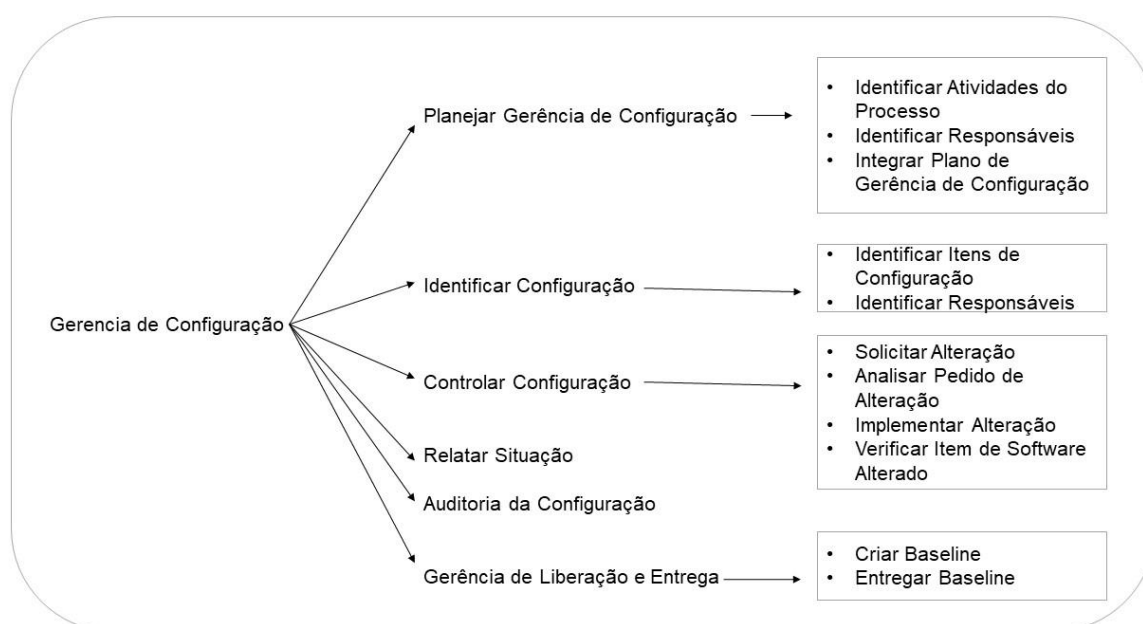
Fonte: Prikladnicki et al. (2004)

Considerando os principais desafios no contexto de Desenvolvimento Distribuído de software, a Gerência de Configuração de Software através dos Sistemas de Controle de Versão ajuda a reduzir os principais problemas que surgem relacionados ao DDS, para que haja controle da grande quantidade de artefatos de software que são produzidos e podem ser alterados a qualquer instante.

## 2.2 GERÊNCIA DE CONFIGURAÇÃO - GC

Ao trabalhar com equipes geograficamente dispersas é necessário que haja o controle dos projetos que são desenvolvidos. A GCS nesse aspecto torna-se um fator importante para controle de sistemas complexos, para evitar a perda do projeto, em virtude dessa grande quantidade de informações que são inseridas e podem ser alteradas em qualquer instante. A abordagem proposta por Figueiredo, Santos e Rocha (2004), descrita na Figura 6 e detalhada posteriormente, para o processo de GCS envolve:

**Figura 6 - Processo de Gerencia de Configuração**



Fonte: Figueiredo, Santos e Rocha (2004, pg.3)

- 1- Planejamento da Gerência de Configuração de Software - as subatividades consistem na identificação das atividades do processo, tendo em vista que, a GCS deve ser identificada levando em consideração as características da organização. Identificação de responsáveis, onde serão identificadas as pessoas responsáveis pela avaliação e aprovação dos pedidos solicitados. E integração do plano de Gerência de

Configuração de Software onde o plano deverá ser atualizado com as atividades previstas para a GCS.

- 2- Identificação de configuração - esta etapa consiste basicamente nos itens que ficarão sobre a Gerência do Controle de Configuração (GCC). Suas subtarefas consistem na identificação dos itens de configuração, na qual o gerente do projeto deve analisar todo o projeto destacando o que deve ou não ter suas modificações controladas. E identificar os responsáveis, em que o gerente pode ser capaz de identificar os responsáveis por cada item de configuração.
- 3- Controlar configuração - possui como objetivo manter e controlar tudo o que foi modificado nos itens de configuração ao longo do processo de desenvolvimento, dentro dessa atividade, qualquer membro da equipe poderá solicitar alteração do item de configuração, analisar o pedido de alteração, onde ao verificar que o pedido foi enviado para alteração, será analisado aprovando ou rejeitando o projeto.
- 4- Relatar situação - possui como objetivo manter todos os envolvidos informados sobre as modificações ocorridas no software.
- 5- Auditoria da configuração - visa assegurar que as alterações tenham sido implementadas corretamente no processo de modificação de software.
- 6- Gerência de liberação e entrega - consiste no gerenciamento de *baselines*, para cliente ou para outro receptor, onde o responsável pela execução da atividade é o responsável pelo projeto, nessa atividade o gerente pode criar uma *baseline*, onde o gerente verifica as versões do projeto. E a atividade de entrega de *baseline*, em que o gerente verifica o pacote e faz a entrega.

Considerando os principais processos, atualmente, um sistema típico do GCS, busca prover serviços para apoiar as principais áreas, na qual segundo Figueiredo et al (2004) destacam:

- Gerenciar um repositório de componentes no qual há uma necessidade de armazenar os diferentes componentes e todas as versões de software de forma segura.
- Apoiar os engenheiros de software em suas atividades usuais - os produtos da GCS tentam prover aos engenheiros, os objetos certos, nos locais certos.
- Apoiar e controlar o processo - O controle das mudanças é parte integral de um produto da GCS e a tendência atual consiste em estender a capacidade de apoio do processo neste aspecto.

A GCS consiste no conjunto de itens que controlam a evolução das configurações no ciclo de vida de um produto de software, essa atividade envolve o controle dos sistemas de versão, atividade responsável por apoiar as atividades de controle de mudanças e integração contínua de um projeto de sistemas de software.

## **2.3 SISTEMAS DE CONTROLE DE VERSÃO - SCV**

A partir das Seções 2.1 e 2.2 sobre os modelos e finalidades no desenvolvimento de GCS, depreende-se a importância da implantação de SCV para os projetos de desenvolvimento de software.

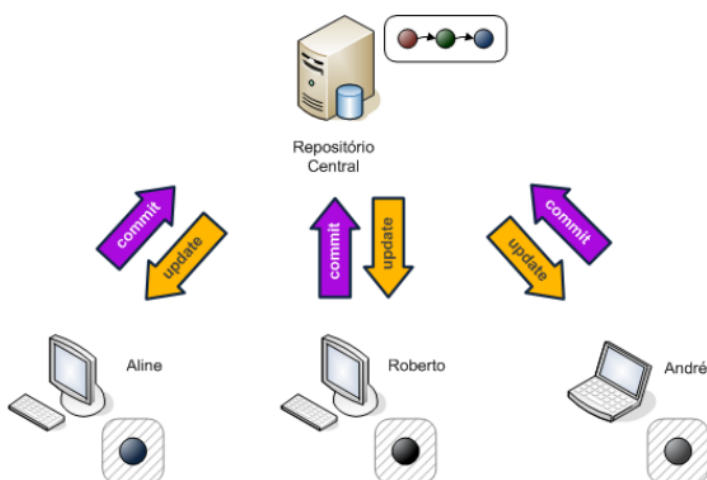
SCV consiste em um sistema que registra as mudanças feitas em artefatos de software ao longo do tempo, de forma que o usuário possa recuperar as versões específicas do projeto no qual está trabalhando. Além disso, o sistema permite reverter arquivos para um estado anterior, ou um projeto desde sua criação, além do acompanhamento das mudanças que ocorreram durante todo o projeto de desenvolvimento, no qual o usuário poderá ter o controle de toda a ação que os colaboradores fazem a partir do momento em que integram o projeto.

No sistema de SCV, segundo Freitas (2010), o repositório irá armazenar todo o histórico de evolução do projeto e o desenvolvedor irá copiar os seus arquivos do repositório para sua área de trabalho, onde ele poderá efetuar as modificações desejadas e salvar os arquivos novamente no repositório. Os SCV podem ser classificados em dois modelos de gerenciamento de repositórios, que segundo Freitas (2010) destacam em: Centralizados e Distribuídos, modelos descritos na Seção 2.3.1.

### 2.3.1 MODELO CENTRALIZADO E DISTRIBUÍDO

Os SCV podem ser classificados em dois modelos de gerenciamento de repositórios sendo estes: Centralizado e Distribuído: Nos modelos Centralizados, segundo Freitas (2010) existe apenas um repositório central e várias cópias de trabalho, neste modelo as operações de envio (*commit*) e atualização (*update*) do repositório acontecem entre cliente e servidor (Figura 7). Quando um commit é realizado, um novo repositório e artefato é criado para quando este passar por modificação. Quando forem atualizados, são enviadas novas modificações contidas no repositório para a área de trabalho.

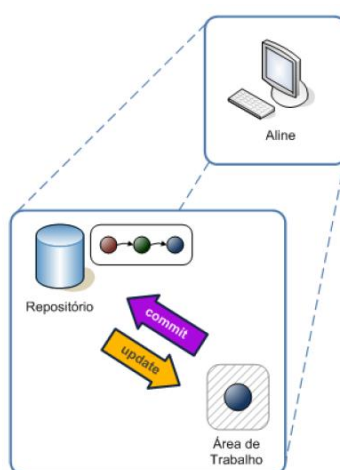
Figura 7 - Modelo Centralizado



Fonte: Freitas (2010)

No modelo distribuído, de acordo com Freitas (2010) existem repositórios autônomos e independentes para cada desenvolvedor, e cada um desses repositórios possui área de trabalho acoplada a ele. Neste momento as operações *commit* e *update* são realizadas localmente. O comando *checkout* é chamado de *clone*, as operações como *commit* e *update* são realizadas localmente além de que é possível realizar a sincronização de repositórios através de *pull* e *push*. A Figura 8 **Erro! Fonte de referência não encontrada.** demonstra como ocorre o gerenciamento desse modelo.

**Figura 8 - Modelo Distribuído**



Fonte: Freitas (2010)

Os dados armazenados nos repositórios de software representam uma fonte significativa de informações e podem ser úteis para compreensão de fatores relacionados ao desenvolvimento e evolução do software (SOARES, 2017). A Seção 2.4 aborda um paradigma de colaboração denominado *Pull Request*, onde um desenvolvedor bifurca do repositório de software, modifica alguns artefatos de seu interesse e, posteriormente, solicita a incorporação de suas modificações no repositório original.

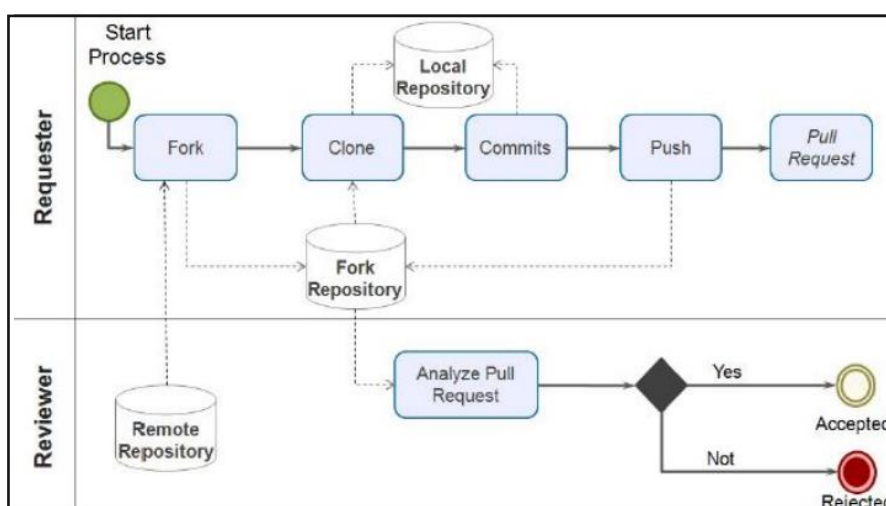


## 2.4 PULL REQUESTS

Um paradigma emergente empregado para a sistematização de contribuições de processo é denominado *Pull Requests*, onde atualmente o desenvolvimento de software de código é inserido em um ambiente colaborativo e distribuído. Segundo Soares (2017), colaboradores externos podem fazer as suas modificações isoladas em artefatos e então solicitar a integração de suas modificações de volta ao repositório central do projeto.

Um *Pull Request* contém um conjunto de alterações em artefatos de software que podem ou não ser incorporadas no projeto principal, recebendo um *Status* de “Aceito” ou “Rejeitado”. A Figura 9 ilustra esse paradigma.

Figura 9 - Processo de Pull



Fonte: Soares (2017)

Inicialmente, o solicitante bifurca seu repositório original, o *fork* nesse caso, é necessário porque os desenvolvedores externos não possuem privilégios sobre o repositório principal. Na segunda fase o solicitante clona o repositório *fork* em seu computador para que a partir de agora ele tenha permissões de leitura e escrita. Na terceira etapa, os artefatos modificados são inicialmente enviados para o repositório recentemente criado através dos *commits*. Na quarta etapa as modificações são enviadas ao repositório *fork* e em seguida o solicitante abre uma solicitação (*Pull Request*) para que as modificações sejam propagadas no repositório central.

Finalmente, os desenvolvedores responsáveis pelo projeto de software decidem se irão integrar ou não as modificações do solicitante.

Embora existam muitas características de *Pull Requests* que sejam conhecidas no momento da abertura, algumas só podem ser observadas após o envio. Algumas características de acordo com Soares (2017), destacam a vida útil do *Pull Request*, o *Status* final da contribuição tais como “aceito” e “rejeitado” e o desenvolvedor responsável por revisá-las as quais são conhecidas somente após o envio para integração do projeto de software.

## 2.5 ACEITAÇÃO DE *PULL REQUESTS* EM PROJETOS OPEN SOURCE

A equipe principal está interessada em receber contribuições para aliviar a carga de trabalho e acelerar o lançamento de novos recursos e correções de erros. A técnica de mineração de dados permite descobrir padrões que fornecem aos gerentes de projeto uma maior compreensão dos recursos de *Pull Requests* enviados ao repositório (SOARES, 2017).

De acordo com Soares (2017), os fatores que podem contribuir na aceitação ou rejeição de projetos em códigos Open Source consistem em:

- 1) Número de *commits* e respostas no pedido *Pull Request*;
- 2) Fatores relativos ao perfil de solicitante, como questões de inexperiência;
- 3) A relação social entre o *requester* e o *reviewer* podem influenciar na aceitação do pedido;
- 4) Fatores como propriedade e localidade de *Pull Request* são importantes as modificações relacionados a localização dos arquivos alterados pelo pedido de *Pull Requests*.

Além de considerar fatores que podem contribuir na aceitação e rejeição do projeto no desenvolvimento baseado em *pull*, a equipe principal está interessada em receber contribuições para aliviar a carga de seus trabalhos e acelerar os recursos e correções de erros (SOARES, 2017). A compreensão de fatores que levam um gerente de projeto aceitar a contribuição pode ajudar os desenvolvedores a orientarem suas ações para que suas solicitações sejam aceitas, além de ajudar os gerentes de equipes a entender a comunidade em torno do projeto de desenvolvimento.

## 2.6 REGRAS DE ASSOCIAÇÃO

As regras de associação caracterizam-se por representar combinações de itens que ocorrem com determinada frequência em uma base de dados (GONÇALVES, 2004). Em uma associação, a regra representa um tipo de padrão de relacionamentos entre itens de dados de um domínio de aplicação que acontece com uma certa frequência no conjunto da base de dados (SOARES et al., 2015).

Uma de suas típicas aplicações, consiste em análise de transações de contas em um determinado lugar, como por exemplo em uma loja de departamentos, ou supermercados (GONÇALVES, 2004). Um exemplo para geração de regras de associação consiste em:  $\{Picanha\} \wedge \{Fraldinha\} \rightarrow \{Toscana\}$ , na qual indica que se o cliente compra picanha e fraldinha, eles tendem a comprar toscana. Sendo assim uma das características mais atrativas em uma regra de associação, é a representação de forma que fique muito fácil para ser entendida (GONÇALVES, 2004).

Uma regra de associação consiste em uma expressão na forma em que  $X \rightarrow Y$  onde  $X \subset I, Y \subset I, X \neq \emptyset, Y \neq \emptyset$  e  $X \cap Y = \emptyset$ . O  $X$  nesta representação é denominado antecedente e o  $Y$  é denominado conseqüente da regra (GONÇALVES, 2004). Geralmente, em grandes bancos de dados que armazenam milhares de itens, como o exemplo do supermercado, deseja-se descobrir associações importantes entre os itens comercializados, tal que a presença de alguns deles em uma transação (compra e vendas) implique na presença de outros fatores em uma mesma transação (ROMÃO et al., 1999).

Para encontrar todas as regras relevantes do tipo  $X(\textit{antecedente}) \rightarrow Y(\textit{consequente})$ , as regras geradas precisam atender a um suporte e confiança mínimos especificados pelo decisor (ROMÃO et al., 1999).

A métrica de suporte é definida pela porcentagem de instâncias que satisfazem as condições do antecedente e do consequente.

$$\textit{Suporte} (\{X\} \rightarrow \{Y\}) = \frac{\textit{Transações contendo } X \textit{ e } Y}{\textit{Número total de transações}}$$

O valor de suporte serve para ajudar a identificar regras que valem a pena considerar para análises posteriores, caso o valor de suporte esteja muito abaixo do esperado, não são obtidas informações suficientes sobre o relacionamento entre os seus itens, logo não serão extraídas nenhuma conclusão sobre a regra que foi estudada (TOWARDS, 2018).

A medição de confiança representa a probabilidade de ocorrência do consequente, dada a ocorrência do antecedente (SOARES et al., 2015).

$$\textit{Confiança} (\{X\} \rightarrow \{Y\}) = \frac{\textit{Transações contendo } X \textit{ e } Y}{\textit{Transações contendo } X}$$

Outra métrica usada para avaliar as regras e identificar quais são interessantes consiste no *lift*, que controla a frequência do consequente ao calcular a probabilidade condicional de ocorrência  $\{Y\}$  dado o  $\{X\}$  (TOWARDS, 2018).

$$\textit{Lift} (\{X\} \rightarrow \{Y\}) = \frac{\textit{Confiança} (X \rightarrow Y)}{\textit{Suporte } Y}$$

Um dos aspectos importantes a serem considerados nos resultados obtidos desta métrica ocorre, se  $\textit{Lift} (X \rightarrow Y) > 1$ , então  $X$  e  $Y$  tornam-se positivamente dependentes. Se  $\textit{Lift} (X \rightarrow Y) < 1$ ,  $X$  e  $Y$  são negativamente dependentes. Esta medida pode variar entre 0 e  $\infty$ , e pode-se chegar a premissa de que: quanto maior for o valor do *lift* de uma determinada regra, mais interessante torna-se a regra, tendo em vista que o  $X$  aumenta o “*lifted*” de  $Y$  em uma taxa maior (GONÇALVES, 2004).

Portanto, ao utilizar regras de associação em determinada base de dados, deve-se observar a frequência em que uma determinada regra pode ocorrer e se ela é ou não relevante, para realizar estudos e descobrir padrões que potencializam ou enfraquecem uma determinada regra. Ao utilizar bases muito grandes para fazer análises, é necessário que sejam utilizados algoritmos rápidos e eficientes e um dos algoritmos mais utilizados para a construção de itens frequentes, consiste no Apriori (VASCONCELOS; CARVALHO,2004).

O algoritmo Apriori, emprega a busca em profundidade e gera conjunto de itens candidatos de  $k$  elementos a partir de conjuntos de itens de  $k - 1$  elementos. Os padrões que não são frequentes na base, são eliminados além disso, toda a base é rastreada e o conjunto de todos os itens frequentes são obtidos a partir dos conjuntos dos itens candidatos (VASCONCELOS; CARVALHO,2004).

O exemplo a seguir demonstra, como funciona a extração de regras de associação para identificar quais são as mais importantes.

A Tabela 1 mostra um exemplo de base de dados com 3 itens e 3 transações, nelas serão avaliados os valores de suporte, confiança e *lift* de uma determinada regra.

**Tabela 1** - Base de dados com 03 transações

<b>Transação</b>	<b>Picanha</b>	<b>Fraldinha</b>	<b>Toscana</b>
<b>1</b>	1	1	0
<b>2</b>	0	1	1
<b>3</b>	0	1	1

Fonte: Elaboração Própria

Dada a base de dados, contendo o registro das transações, o suporte da regra de conjuntos com 2 item (*itemsets*) é exemplificada na Tabela 2.

**Tabela 2** - Cálculo do Suporte

Item	Suporte
Picanha	$1/3 = 33\%$
Fraldinha, Toscana	$2/3 = 66\%$
Toscana	$2/3 = 66\%$

Fonte: Elaboração Própria

Dado os resultados do suporte, infere-se por exemplo a taxa de frequência em que a fraldinha e toscana são vendidos juntos ocorrem em cerca de 66% das transações. A confiança da regra, consiste na probabilidade de ocorrência das transações contendo X e Y sobre a transação contendo X em uma regra.

**Tabela 3** - Cálculo da confiança e *lift* da regra

Item	Suporte	Confiança	<i>Lift</i>
Picanha -> Fraldinha	$1/3 = 33\%$	$1/1 = 100\%$	$1/1 = 1$
Fraldinha -> Toscana	$2/3 = 66\%$	$2/3 = 66\%$	$66/66 = 1$

Fonte: Elaboração Própria

Os resultados gerados descritos na Tabela 3, indicam a confiança em que, todas as vezes em que uma pessoa compra picanha então ela tem 100% de chance de comprar fraldinha. A mesma regra mostrou-se com o *lift* igual a 1, tornando esta regra independente.

A ferramenta utilizada para agregar os algoritmos provenientes de diferentes sub áreas consiste no WEKA<sup>1</sup> que é uma coleção de algoritmos de aprendizado de

---

<sup>1</sup> [cs.waikato.ac.nz/ml/weka](http://cs.waikato.ac.nz/ml/weka)

máquina para tarefas concernentes a mineração de dados. Nesta ferramenta, estão contidas as ferramentas para classificação, regressão, clustering, mineração de regras de associação e visualização de dados (WEKA, 2019).

O plugin de extensão que foi utilizado para adicionar recursos para a visualização, filtragem e análise de regras de associação geradas a partir da ferramenta de mineração de dados WEKA consistiu no WEKAPAR<sup>2</sup>, cujo objetivo é fornecer ao usuário, um ambiente onde as regras mais interessantes possam ser selecionadas, dado um conjunto de parâmetros definidos pelo usuário (WEKAPAR, 2016).

## 2.7 CONSIDERAÇÕES FINAIS SOBRE O CAPÍTULO

Considerando os principais tópicos abordados no decorrer do Capítulo 2, depreende-se os principais fatores que dão embasamento teórico para a realização do estudo de caso abordado no Capítulo 3.

O tópico abordado ao longo do capítulo tem como base justificar o principal objetivo abordado neste trabalho que consiste na influência da localização geográfica do *requester* nos cenários de aceitação e tempo de vida do *Pull Request* e o relacionamento com fatores físicos das contribuições. Para compreensão destes fatores é necessário que sejam abordadas questões como localização geográfica e como ocorre o gerenciamento das solicitações que são enviadas pelos desenvolvedores em qualquer lugar do mundo.

---

<sup>2</sup> <https://github.com/gems-uff/wekapar/wiki>

### **3 RESULTADOS E DISCUSSÕES**

Com base nos principais conceitos abordados no Capítulo 2, neste capítulo será abordado o estudo de caso realizado, com o objetivo de investigar a hipótese que foi levantada nos capítulos anteriores.

A Seção 3.1 tem como objetivo responder, se a localização geográfica e o *lifetime* influenciam na aceitação/rejeição de *Pull Requests*.

A Seção 3.2 responde a hipótese de que, dado um país de origem de um *Pull Request* influencia na aceitação/rejeição dos projetos, quais outras características estão presentes nestes *Pull Requests*.

A Seção 3.3 tem como objetivo responder a terceira questão da hipótese levantada, indicando quais são os fatores que potencializam ou enfraquecem a influência do país.

#### **3.1 INFLUÊNCIA DA LOCALIZAÇÃO GEOGRÁFICA DO REQUESTER NA ACEITAÇÃO *PULL REQUESTS***

Para fins de verificação das questões de pesquisa levantadas neste trabalho que consiste em verificar se a localização geográfica e o *lifetime* influenciam na aceitação/rejeição de *Pull Requests*, quais características estão presentes nestes *Pull Requests* e quais são os fatores que potencializam ou enfraquecem a influência do



país, foi explorada uma base de dados com o objetivo de compreender se a localização geográfica tem influência na aceitação e tempo de vida de *pull request*.

Para a realização deste estudo de caso, reutilizou-se o *dataset* estudado na tese de doutorado de Soares (2017). Após pré-processamento a base resultou na junção de 26 projetos (Descritos no APÊNDICE A), 7553 instâncias (*Pull Requests*), 61 atributos, contendo contribuições feitas por desenvolvedores de 88 países diferentes (APÊNDICE B). Dos 61 atributos que a base possui, foram considerados para análise deste estudo apenas 7 atributos (Quadro 1), que consistem nos principais fatores que contextualizam e respondem as hipóteses levantadas neste trabalho.

**Quadro 1** - Atributos para a extração das regras

<b>Atributos</b>	<b>Descrição dos Atributos</b>
<i>Country</i>	Países nos quais os <i>requesters</i> que fazem as modificações dos projetos estão localizados.
<i>lifetimeMinutes</i>	Tempo de vida do <i>Pull Request</i> , onde "Very Short" = até 1 dia; "Short" = entre 1 a 3 dias; "Medium" = 3 a 7 dias; "Lengthy" = 7 a 10 dias; e "Very Lengthy" = mais de 10 dias.
<i>commitsPull</i>	Número de <i>commits</i> do <i>Pull Request</i> , onde os valores dos rótulos distribuídos consistem em: "1 commit"; "some commits": 2 a 4 <i>commits</i> ; e "many commits": maior que 4 <i>commits</i> .
<i>changedFiles</i>	O número de arquivos adicionados, removidos ou editados. Os rótulos aplicados consistem em: "1 File"; "Some Files": 2 a 4 arquivos; e "Many Files": maior que 4 arquivos.
<i>Status</i>	O <i>Status</i> final do <i>Pull Request</i> : na qual podem ser "merged" e "closed".

<b>Atributos</b>	<b>Descrição dos Atributos</b>
<i>Language</i>	A linguagem de programação dominante no projeto que está sendo analisado.
<i>TotalLines</i>	O número de linhas de código manipuladas distribuídos em: “1 <i>line</i> ”; “ <i>Some Lines</i> ”: 2 a 20 linhas; “ <i>Many Lines</i> ”: maior que 20 linhas

Fonte: Adaptada de Soares (2017).

Inicialmente, alguns atributos da base eram dados numéricos e para a realização da análise foi necessário discretizar estes atributos. Para realizar a discretização foram organizados a base em uma planilha que consistiu no Excel<sup>3</sup>, no qual foram criados novos atributos com rótulos especificados no Quadro 1 e após isso foram extraídas as regras de associação para posterior análise dos dados.

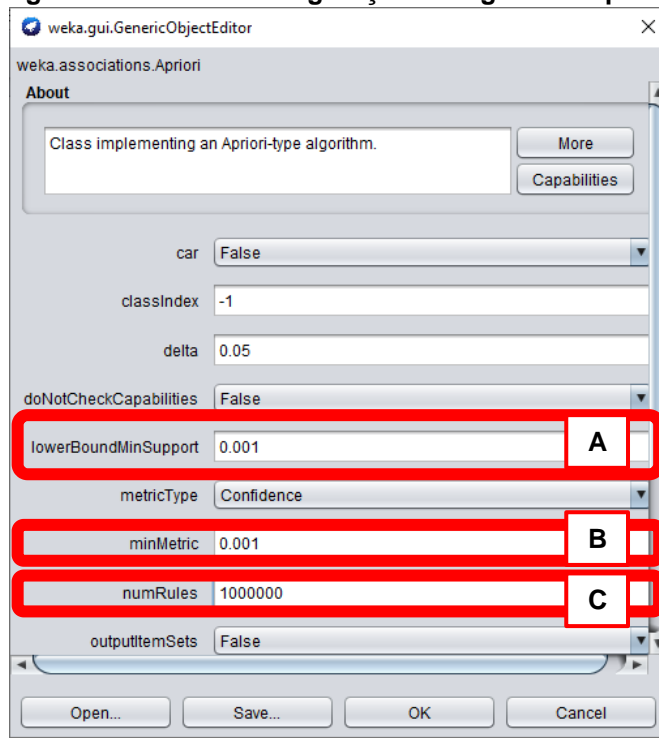
A ferramenta utilizada para realizar a extração das regras consistiu no WEKA, e o algoritmo para mineração dos dados o Apriori<sup>4</sup>. Esta ferramenta, demonstrada na Figura 10. Inicialmente definiu-se os valores de:

- Suporte mínimo: configurado para 0,001 (A);
- Métrica mínima: configurada para 0,001 (B);
- Número de regras: configurado em 1000000 de regras (C).

<sup>3</sup> <https://products.office.com/pt-br/excel>

<sup>4</sup> <http://weka.sourceforge.net/doc.dev/weka/associations/Apriori.html>

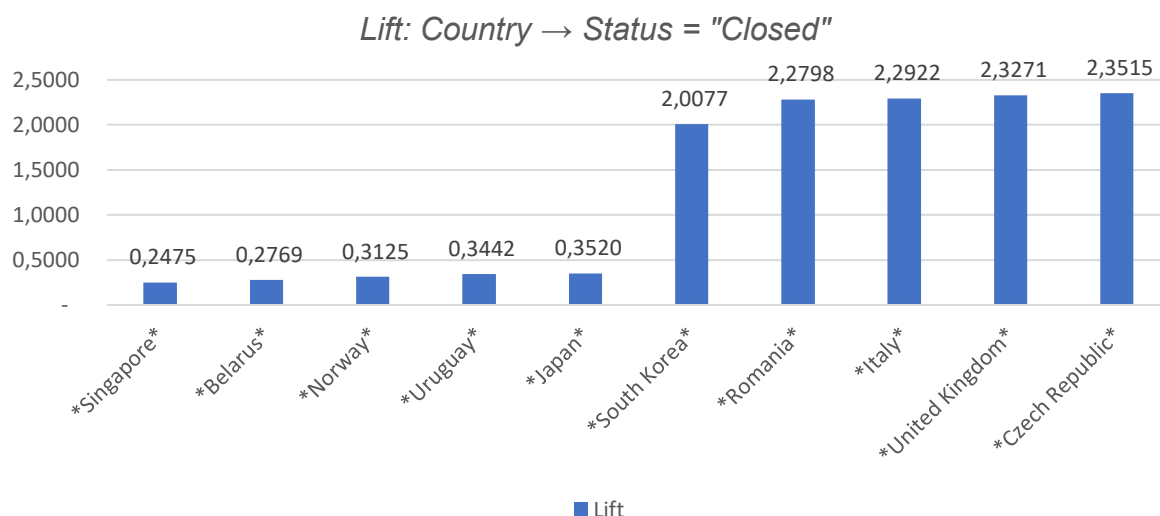
**Figura 10 - Tela de configuração do algoritmo Apriori**



Fonte: Elaboração Própria

As regras que foram geradas para visualizar o nível de influência da localização geográfica consideraram os seguintes atributos: *Country*, *Status*, *lifetimeMinutes*.

Para análise dos resultados, utilizou-se a regra que envolve os atributos *Country* → *Status*. Quando o *Status* for “closed”, 42 desenvolvedores oriundos de qualquer país do mundo têm probabilidade de quando submeterem *Pull Requests* sejam fechados. Foram analisados o *lifts* destas regras, uma vez que indicam o quão frequente torna-se o *Status* = “closed” dado que um determinado desenvolvedor oriundo de determinado país submete um projeto. A Figura 11 mostra o gráfico com os países que possuem as maiores e menores chances de serem fechadas.

**Figura 11 – Lift da regra do tipo: Country → Status = "closed"**

Fonte: Elaboração Própria

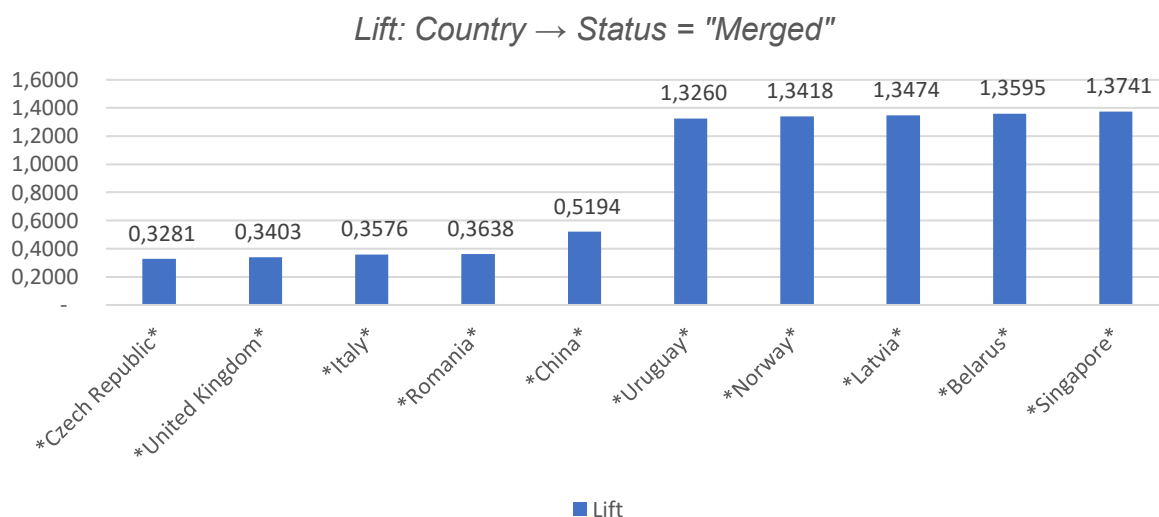
A Figura 11 evidencia que os desenvolvedores oriundos de países como Singapura, Bielorrússia e Noruega possuem chances de rejeição reduzidas, enquanto Itália, Reino Unido e República Checa aumentam as chances de rejeição. De acordo com Soares (2017), as regras de associação possibilitam perceber não apenas o aumento/diminuição das chances de rejeição, mas quantificar a variação que ocorre nestas regras.

Analisando o *lift* dos países na regra *Country → Status = "closed"* é possível analisar que os desenvolvedores provenientes de Singapura, possuem 75,25% menos chances de serem fechadas quando submetem *Pull Request* (*lift* = 0,2475). Quando são oriundos de países como Bielorrússia e Noruega, possuem respectivamente 72,31% e 68,75% menos chances de serem fechadas (*lift*= 0,2769, *lift*=0,3125).

Sob outra perspectiva, um *Pull Request* oriundo da Itália tem uma chance 129% maior de ser rejeitado (*lift*= 2,2922). Da mesma forma, quando são submetidos no Reino Unido e República Checa detêm-se chances maiores em serem rejeitados cerca de 132% e 135%, respectivamente (*lift* = 2,3271, *lift* = 2,3515).

A Figura 12 mostra os *lifts* das regras do tipo *Country* → *Status* = “merged”, nesta regra foram filtrados 47 países que possuem os maiores e menores índices de serem aceitas quando submetem *Pull Request*.

**Figura 12 – Lift da regra do tipo: *Country* → *Status* = “merged”**



Fonte: Elaboração Própria

Analisando a regra é possível identificar que *Pull Requests* cujos desenvolvedores são oriundos da República Checa têm 67% menos chances de aceitação quando submetem seus projetos (*lift*=0,3281). Da mesma forma, quando são provenientes dos países do Reino Unido e Itália, possuem respectivamente 65,97%, 64,24% menos chances de aceitação (*lift*= 0,3403, *lift* = 0,3576).

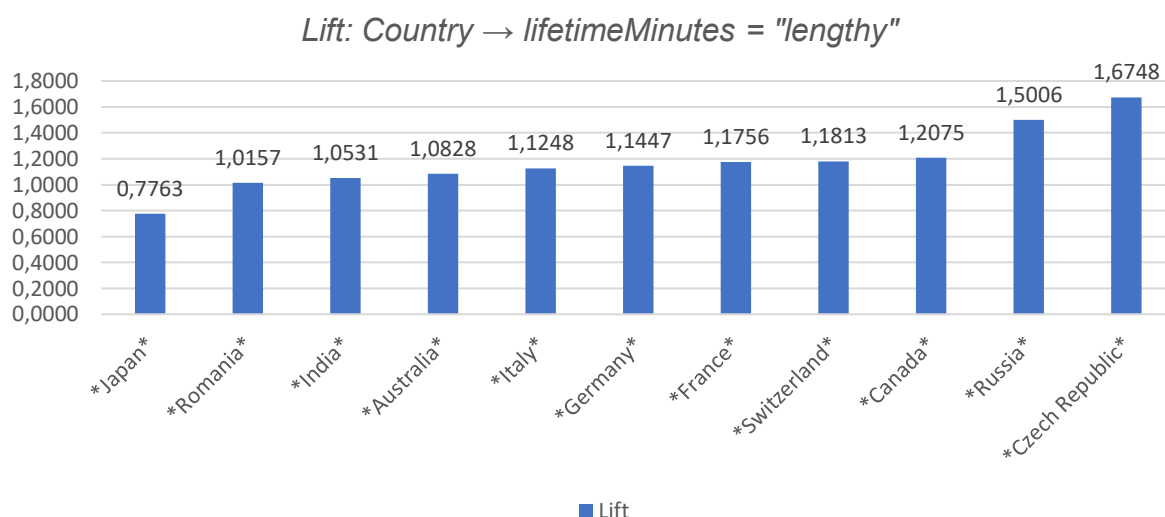
Por outro lado, quando os desenvolvedores são originários da Singapura, tem chance 37% maior de serem aceitas (*lift* =1,3741), em Bielorrússia e Letônia, tem chances maiores de serem aceitos em, respectivamente, 35% e 34% (*lift*=1,3595, *lift* = 1,3474).

Observando que um país tem influência no *Status* de aceitação e rejeição de um projeto de colaboração, a próxima regra consistiu em *Country* → *lifetimeMinutes*, com o objetivo de verificar se o tempo de vida do *Pull Request* sofre influência da

localização geográfica dos desenvolvedores. Ao aplicar este filtro foram retornadas 151 regras.

Quando o tempo de análise de *Pull Requests* for demorado (*Country* → *lifetimeMinutes* = "*Lengthy*"), 11 países apresentam algum tipo de influência quando submetem *Pull Request*. A Figura 13, evidencia estes países que aumentam ou reduzem as chances de serem analisados de forma demorada.

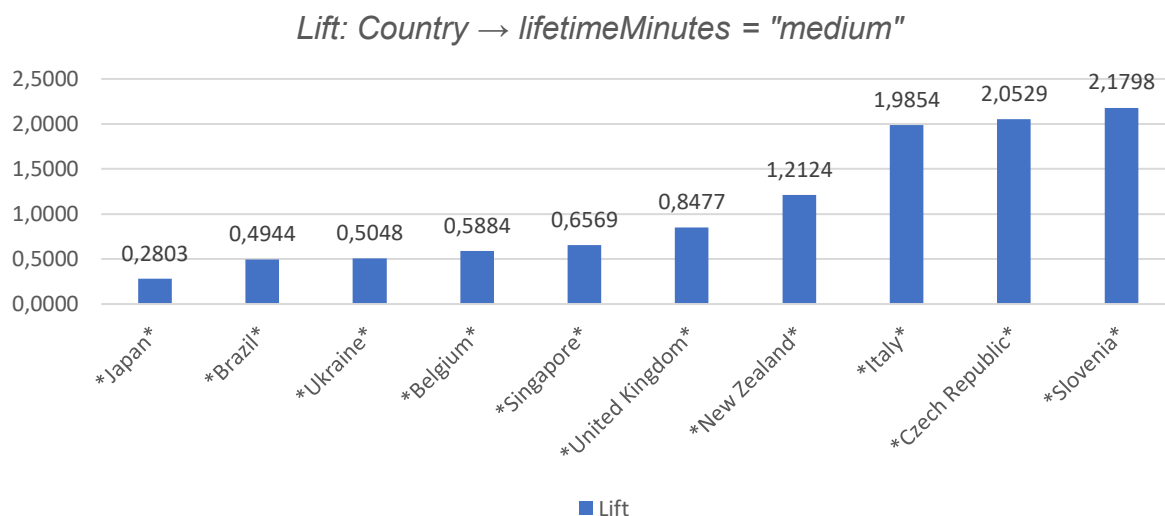
**Figura 13 – Lift da regra do tipo: *Country* → *lifetimeMinutes* = "*lengthy*"**



Fonte: Elaboração Própria

A Figura 13, mostra os valores da métrica *lift* para as regras do tipo *Country* → *lifetimeMinutes* = "*Lengthy*" e permitem a observação que, desenvolvedores oriundos de países como Itália e República Checa possuem dependências positivas em 12% (*lift*=1,1248) e 67% (*lift*=1,6748), ou seja, maiores chances de tempo de vida mais longo.

Quando o *lifetime* da regra for em tempo médio (*lifetimeMinutes* = "*Medium*"), 25 países quando os desenvolvedores submetem seus projetos são analisados entre 3 a 7 dias. A Figura 14 ilustra o cenário com os principais países.

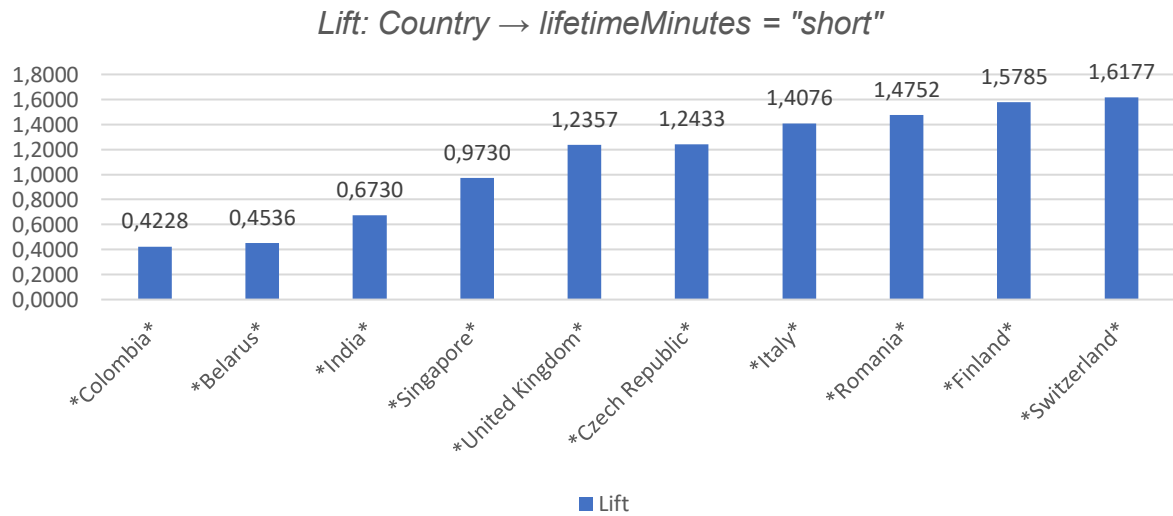
**Figura 14** – *Lift* da regra do tipo: *Country* → *lifetimeMinutes* = "medium"

Fonte: Elaboração Própria

Na Figura 14 é possível identificar que os desenvolvedores provenientes de países como Eslovênia, República Checa e Itália que possuem os maiores *lifts*, são analisados entre 3 a 7 dias.

Por outro lado, considerando países que possuem menores *lifts*, indicando dependências negativas, observa-se que desenvolvedores oriundos do Japão, Brasil e Ucrânia possuem respectivamente 71,97%, 50,56% e 49,52% menores chances de serem analisados em tempo de vida médio.

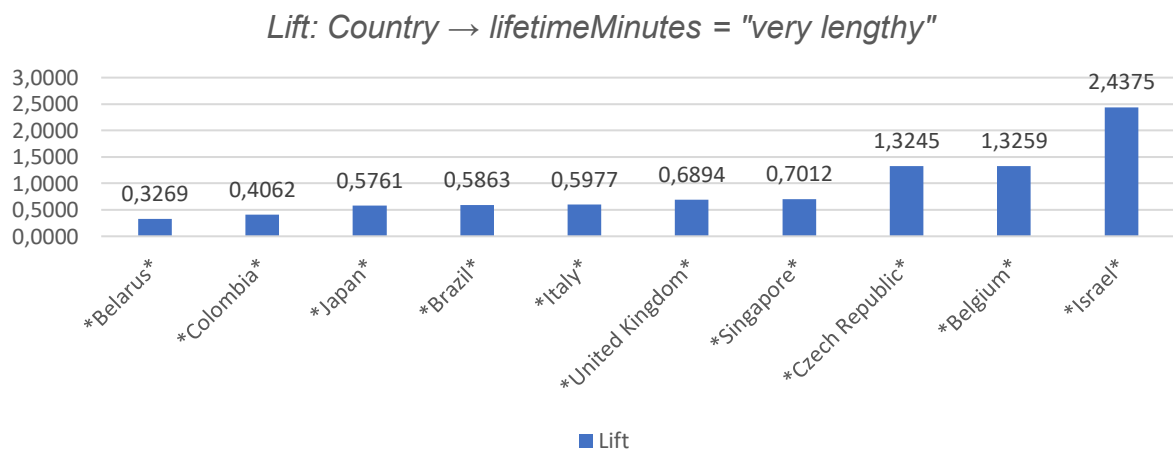
Na Figura 15, observa-se que quando o *lifetime* do *Pull Request* for analisado entre 1 a 3 dias (*Country* → *lifetimeMinutes* = "Short"), países como Singapura, por exemplo, possui dependência negativa (*lift* < 1), indicando uma chance 2,7% menor de ser analisado em tempo rápido. Em Bielorrússia a dependência do projeto torna-se negativa quando o *lifetime* = "Short" cerca de 54,64% (*lift* = 0,4536).

**Figura 15 – Lift da regra do tipo: Country → lifetimeMinutes = "short"**

Fonte: Elaboração Própria

Países como Itália, Reino Unido e República Checa, possuem dependências positivas, quando o tempo de vida de um *Pull Request* for analisado em até 3 dias ( $lift=1,40$ ,  $lift=1,23$ ,  $lift=1,24$ ). A Figura 15 mostra as variações do *lift* quando o *lifetimeMinutes* = "Short".

Quando o *lifetime* do *Pull Request* for muito longo (*Country* → *lifetimeMinutes* = "Very Lengthy"), as dependências variam possuindo, dependências negativas e positivas (Figura 16).

**Figura 16 – Lift da regra do tipo: Country → lifetimeMinutes = "very lengthy"**

Fonte: Elaboração Própria

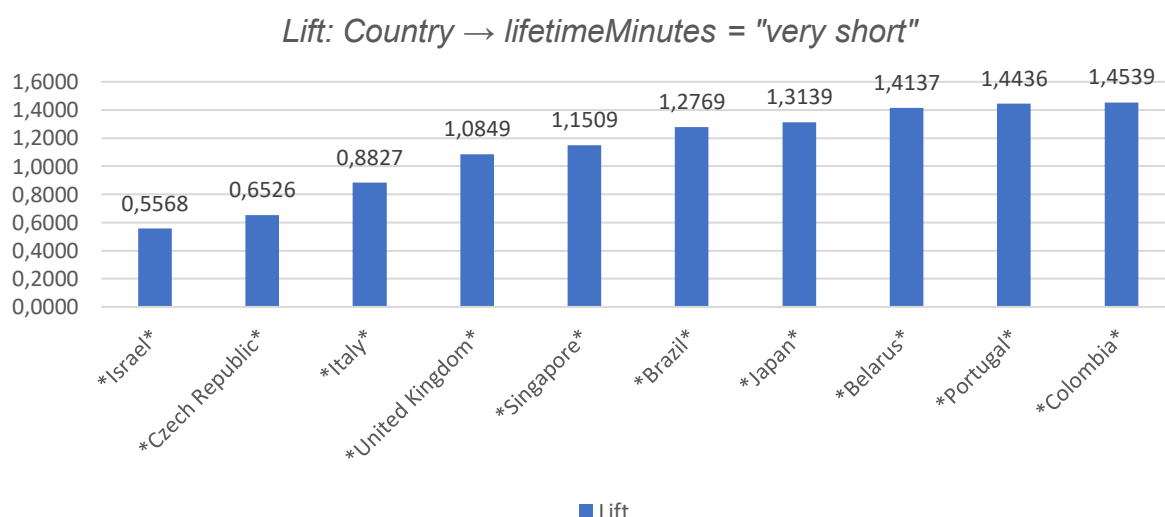


Ao visualizar a Figura 16 observa-se que desenvolvedores oriundos de Israel, possuem 143% maiores chances de serem analisados em tempo de vida muito longo (maior que 10 dias) indicando dependência positiva nesta regra, assim ocorrem nos países da Bélgica e República Checa que possuem dependências positivas ( $lift > 1$ ), indicando que quando os desenvolvedores forem provenientes destes países eles possuem respectivamente 32,59% e 32,45% maiores chances de serem analisados em tempo de vida muito longo.

Em contrapartida, desenvolvedores oriundos de Bielorrússia, possuem dependências negativas na regra, indicando que a ocorrência deste país está negativamente correlacionada com o  $lifetime = \text{"Very Lengthy"}$  ( $lift = 0,3269$ ), e na Itália possui 40,23% menores chances de serem analisados em tempo de vida muito longo.

Quando a regra consistir em  $Country \rightarrow lifetimeMinutes = \text{"very Short"}$  países como Letônia que possui chance 34% maior de serem aceitos ( $lift = 1,3474$ ), possui 55% ( $lift = 1,5591$ ) do  $lifetime$  da regra ser muito rápido. A mesma regra vale para os países da Singapura e Bielorrússia que possuem dependência positiva de 15% e 41% (Figura 17).

**Figura 17 – Lift da regra do tipo:  $Country \rightarrow lifetimeMinutes = \text{"very short"}$**



Fonte: Elaboração Própria

Países como Itália e República Checa apresentam dependências negativas quando o projeto for analisado em até 24 horas, cerca de 11,73% e 34,74% ( $lift = 0,8827$ ,  $lift = 0,6526$ ). Desenvolvedores oriundos do Reino Unido possuem influência positiva em 8,49% quando o tempo de análise do *Pull Request* for muito rápido ( $lift = 1,0849$ ).

### 3.2 CARACTERÍSTICAS DE ACEITAÇÃO DE *PULL REQUESTS*

Considerando a hipótese de que a localização geográfica pode influenciar no *status* e *lifetime* conforme abordado na Seção 3.1, demonstrando os valores que potencializam ou enfraquecem as regras, esta Seção tem como objetivo apontar algumas características que justifiquem os achados da Seção 3.1. Como por exemplo um país que possui grandes chances de rejeição replica padrões ruins ou quando ele possui grandes chances de aceitação são replicados padrões bons (exemplos: número de linhas modificadas, tempo de vida).

Logo, para categorizar a regra afim de justificar que país influencia no *status* e *lifetime* foram analisadas regras em que *Country* é o antecedente e alguns fatores físicos o consequente, a saber: número de arquivos, tempo de vida e número de linhas modificadas (*changedFiles*, *lifetimeMinutes* e *totalLines*).

Com o filtro *Country* → *changedFiles* foram retornadas 121 regras. Os países em que são modificados apenas 1 arquivo que possuem maiores probabilidades de serem aceitas são exibidas na Tabela 4.

**Tabela 4** – Países com maiores probabilidades de aceitação quando  
*Country* → *changedFiles* = “1 File”.

#	Antecedente	Consequente	Sup.	Conf.	Lift
1	*Singapore*	1 File	1,11%	57,53%	1,2945
2	*Belarus*	1 File	1,11%	48,28%	1,0862
3	*Brazil*	1 File	1,13%	43,81%	0,9858
4	*United States*	1 File	0,3%	53,49%	1,2034

Fonte: Elaboração Própria

Ao visualizar a regra, observa-se que *Pull Requests* cujos desenvolvedores são oriundos de Singapura, possuem cerca de 37% mais chances em serem aceitos, e em 57,53% das vezes o desenvolvedor modifica apenas um arquivo. O mesmo fator, aplica-se quando o país for Bielorrússia, onde 48,28% das vezes em que um *Pull Request* é submetido de lá apenas um arquivo é modificado. Países como Brasil e Estados Unidos, observa-se que os valores de confiança ao modificarem apenas um arquivo consistem em 43,81% e 53,49%.

Considerando os países que possuem menores chances de serem aceitas, observa-se que os valores da confiança variam entre 32% a 42% conforme a Tabela 5.

**Tabela 5** - Países com menores probabilidades de aceitação quando  
*Country* → *changedFiles* = “1 File”.

#	Antecedente	Consequente	Sup.	Conf.	Lift
1	*Italy*	1 File	1,34%	32,17%	0,7237
2	*United Kingdom*	1 File	1,11%	42,42%	0,9545
3	*Czech Republic*	1 File	0,69%	35,62%	0,8013

Fonte: Elaboração Própria

Considerando a Tabela 5, observa-se que na Itália em 32,17% das vezes quando são aceitos o *Pull Requests* são modificados apenas 1 arquivo. Quando apenas um arquivo é modificado no Reino Unido, cerca de 42,42% das vezes têm menores chances de serem aceitos. Em 35,62% das vezes em que um arquivo é modificado na República Checa, tende a ser rejeitado.

Quando *Pull Requests* oriundos de países que possuem maiores probabilidades de serem aceitos e conjuntamente existem modificações em muitos arquivos, tendem a não possuir uma influência forte positivamente, conforme a Tabela 6.

**Tabela 6** - Países com maiores probabilidades de aceitação quando  
*Country* → *changedFiles* = “many files”.

#	Antecedente	Consequente	Sup.	Conf.	Lift
1	*Singapore*	<i>Many Files</i>	0,21%	10,96%	0,5249
2	*Belarus*	<i>Many Files</i>	0,29%	12,64%	0,6056

Fonte: Elaboração Própria

Em contrapartida, considerando os países que possuem menores probabilidades de serem aceitos conforme abordado na Seção 3.1, os resultados são atenuados indicando que quando o desenvolvedor modificar muitos arquivos talvez não seja uma regra muito interessante, tendo em vista que seus projetos tendem a ter menores probabilidades de aceitação. A Tabela 7, descrevem as variações quando ocorre esta regra.

**Tabela 7** - Países com menores probabilidades de aceitação quando  
*Country* → *changedFiles* = “*Many Files*”.

#	Antecedente	Consequente	Sup.	Conf.	Lift
1	*Italy*	<i>Many Files</i>	1,50%	35,99%	1,7236
2	*United Kingdom*	<i>Many Files</i>	0,78%	29,80%	1,4272
3	*Czech Republic*	<i>Many Files</i>	0,58%	30,14%	1,4434

Fonte: Elaboração Própria

Analisando esta regra observa-se, quando os arquivos são submetidos na Itália, em 35,99% das vezes são modificados muitos arquivos. As mesmas regras colaboram quando os países forem do Reino Unido e República Checa onde são modificados cerca de 29,80 a 30,14% das vezes muitos arquivos.

A Tabela 8, mostra a regra *Country* → *changedFiles* = “*Some Files*”. Considerando os países que possuem maiores probabilidades de serem aceitos, observa-se que em 31,51% de vezes Singapura modifica alguns arquivos. Bielorrússia realiza modificações em alguns arquivos em 39,08% de vezes.

**Tabela 8** - Países com maiores probabilidades de aceitação quando  
*Country* → *changedFiles* = “*Some Files*”.

#	Antecedente	Consequente	Sup.	Conf.	Lift
1	*Singapore*	<i>Some Files</i>	0,61%	31,51%	0,9086
2	*Belarus*	<i>Some Files</i>	0,90%	39,08%	1,1271

Fonte: Elaboração Própria

Considerando países com menores probabilidades de aceitação, conforme a Tabela 9, observa-se que a confiança da regra que ocorre possui variação entre 27 a 34,25% das vezes.

**Tabela 9** - Países com menores probabilidades de aceitação quando  
*Country* → *changedFiles* = "Some Files".

#	Antecedente	Consequente	Sup.	Conf.	Lift
1	*Italy*	<i>Some Files</i>	1,32%	31,85%	0,9184
2	*United Kingdom*	<i>Some Files</i>	0,73%	27,78%	0,8011
3	*Czech Republic*	<i>Some Files</i>	0,66%	34,25%	0,9876

Fonte: Elaboração Própria

Quando um contribuidor é oriundo da Itália em 31,85% de chances eles modificam alguns arquivos. Considerando que um contribuidor seja do Reino Unido e República Checa, em 27,78% e 34,25% de chances eles modificam alguns arquivos.

A próxima regra analisada consistiu em números de linhas modificadas através do filtro: *Country* → *totalLines*, no qual foram retornadas 101 regras. As principais características consistiram em: "Some Lines": 2 a 20 linhas e "Many Lines": maiores que 20 linhas.

*Pull Requests* oriundos de alguns países que possuem maiores chances de aceitação, ao modificarem muitas linhas, a confiança gerada pela regra varia entre 39 a 42,53% conforme a Tabela 10.

**Tabela 10** - Países com maiores probabilidades de aceitação quando  
*Country* → *totalLines* = "Many Lines"

#	Antecedente	Consequente	Sup.	Conf.	Lift
1	*Singapore*	<i>Many Lines</i>	0,75%	39,04%	0,8342
2	*Belarus*	<i>Many Lines</i>	0,98%	42,53%	0,9087

Fonte: Elaboração Própria

Analisando esta regra percebe-se que em Singapura, são manipuladas muitas linhas em 39,04% das vezes. Em Bielorrússia, 42,53% das vezes são manipuladas muitas linhas.

Considerando países que possuem menores chances de aceitação quando modificam muitas linhas, observa-se que o fator de confiança aumenta, em Singapura 58,28% das vezes modificam muitas linhas, Reino Unido possui confiança de 51,01% e República Checa cerca de 64,38% (Tabela 11).

**Tabela 11** - Países que possuem menores chances de aceitação quando  
*Country → totalLines = "Many Lines"*

#	Antecedente	Consequente	Sup.	Conf.	Lift
1	*Italy*	<i>Many Lines</i>	2,42%	58,28%	1,2452
2	*United Kingdom*	<i>Many Lines</i>	1,34%	51,01%	1,0899
3	*Czech Republic*	<i>Many Lines</i>	1,24%	64,38%	1,3756

Fonte: Elaboração Própria

Considerando os casos em que o desenvolvedor modifica entre 2 a 20 linhas em um projeto de desenvolvimento, observa-se que os países que possuem maiores chances de aceitação em relação a quando modificam muitas linhas, tem um aumento significativo da confiança indicando que tornam-se mais sensato um desenvolvedor modificar apenas algumas linhas ao invés de muitas. A Tabela 12, mostra a variação da confiança nos países em que são modificadas algumas linhas.

**Tabela 12** - Países que possuem maiores chances de aceitação quando  
*Country → totalLines = "Some Lines"*

#	Antecedente	Consequente	Sup.	Conf.	Lift
1	*Singapore*	<i>Some lines</i>	1,11%	57,53%	1,1382
2	*Belarus*	<i>Some lines</i>	1,27%	55,17%	1,0915

Fonte: Elaboração Própria

Em Singapura, 57,53% dos desenvolvedores quando submetem seus projetos modificam apenas algumas linhas. Em Bielorrússia, 55,17% dos desenvolvedores modificam algumas linhas antes de enviarem suas solicitações para incorporar ao projeto principal.

Em países que possuem menores chances de aceitação, observa-se que a confiança da regra tende a cair comparada quando o desenvolvedor modifica muitas linhas, indicando menores chances de serem rejeitados. A Tabela 13, mostra a confiança gerada pela regra em países que possuem menores probabilidades em serem aceitas.

**Tabela 13** - Países que possuem menores chances de aceitação quando  
Country → totalLines = "Some Lines"

#	Antecedente	Consequente	Sup.	Conf.	Lift
1	*Italy*	<i>Some Lines</i>	1,63%	39,17%	0,7749
2	*United Kingdom*	<i>Some Lines</i>	1,23%	46,97%	0,9292
3	*Czech Republic*	<i>Some Lines</i>	0,66%	34,25%	0,6775

Fonte: Elaboração Própria

Na Itália, 39,17% dos desenvolvedores modificam algumas linhas do projeto para realizar a integração. No Reino Unido, observa-se que este valor aumenta, quando um usuário modifica algumas linhas em 46,97%, Na República Checa, quando são modificados algumas linhas, a confiança gerada é em torno de 34,25% de chances. A confiança gerada pela regra, indica a validade da regra acontecer quando as transações acontecerem.

### 3.3 FATORES DE INFLUÊNCIA NO PAÍS

Ao considerar que país têm influência na aceitação e tempo de vida de *Pull Requests* abordados da Seção 3.1, e ao identificar as principais características que atenua nesta aceitação (Seção 3.2) considerando alguns fatores como número de linhas, tempo de vida e número de arquivos modificados no projeto, observa-se que estes aspectos corroboram com linhas de pesquisas encontradas em arcabouços teóricos de Soares (2017), na qual justifica que as combinações de características, individualmente importantes, aumentam as chances de aceitação em até 70%.



Considerando estes principais aspectos, esta seção tem como objetivo correlacionar os fatores de aceitação de rejeição de *Pull Requests* indicando as possíveis regras que ao serem geradas aumentem as chances de aceitação, considerando um fator que é abordado no trabalho de Soares (2017), em que todos têm interesse que suas solicitações sejam aceitas.

A regra plotada para a confirmação desta hipótese consistiu em:  $Country \wedge Factors \rightarrow Status$ , na qual os fatores que foram analisados embasaram-se em número de arquivos modificados, tempo de vida e número de linhas modificadas (*changedFiles*, *lifetimeMinutes* e *totalLines*).

Inicialmente a regra filtrada consistiu em  $Country \wedge changedFiles \rightarrow Status$ . A partir deste filtro foram retornadas cerca de 169 regras. Em Singapura que possui maior chance de aceitação quando submete um projeto ( $lift = 1,3741$ ), será aceita em 97,83% das vezes quando forem modificados entre 2 a 4 arquivos. Quando forem modificadas 1 arquivo serão aceitas em 91,67% das vezes, e quando houver modificações em muitos arquivos, serão aceitas em 75% das vezes (Tabela 14).

**Tabela 14** - Regras de associação relevantes para aceitação quando houverem modificações de arquivos em Singapura

#	Antecedente	Consequente	Sup.	Conf.	Lift
1	*Singapore* ^ 1 File	Merged	1,02%	91,67%	1,3724
2	*Singapore* ^ Many Files	Merged	0,16%	75,00%	1,1228
3	*Singapore* ^ Some Files	Merged	0,60%	97,83%	1,4646

Fonte: Elaboração Própria

Quando o país de submissão do *Pull Request* for Bielorrússia que é o segundo país com maiores chances de aceitação ( $lift = 1,3595$ ), serão aceitas em 97,06% das vezes quando forem alterados alguns arquivos. Quando modificados muitos arquivos serão aceitos 72,73% e quando houver modificação em apenas um arquivo cerca de 90,48% das vezes serão aceitas (Tabela 15). Um fator importante que deve ser considerado é que quando são modificados alguns arquivos, há um aumento na variação do *lift*, indicando que essa característica se torna um fator de influência para que os projetos sejam aceitos.

**Tabela 15** - Regras de associação relevantes para aceitação quando houverem modificações de arquivos em Belarus

#	Antecedente	Consequente	Sup.	Conf.	Lift
1	* Belarus* ^ 1 File	<i>Merged</i>	1,01%	90,48%	1,3545
2	* Belarus* ^ Many Files	<i>Merged</i>	0,21%	72,73%	1,0888
3	* Belarus* ^ Some Files	<i>Merged</i>	0,87%	97,06%	1,4531

Fonte: Elaboração Própria

Na Itália quando são modificados muitos arquivos, em 87,61% das vezes são fechadas. Além disso, este país detém de 163% de chances maiores em serem fechadas comparadas apenas ao *lift* do país (*lift* = 2,3515), indicando que este fator possui influência quando são modificados muitos arquivos (Tabela 16).

**Tabela 16** - Regras de associação relevantes para rejeição quando houverem modificações de arquivos em Itália

#	Antecedente	Consequente	Sup.	Conf.	Lift
1	*Italy*^ 1 File	<i>Closed</i>	0,83%	62,38%	1,8785
2	*Italy*^ Many Files	<i>Closed</i>	1,31%	87,61%	2,6384
3	*Italy*^ Some Files	<i>Closed</i>	1,02%	77,00%	2,3189

Fonte: Elaboração Própria

No Reino Unido possuem cerca de 132% maiores chances de rejeição, potencializam este *Status* quando modificam entre alguns e muitos arquivos (cerca de 162% e 139%). Quando são modificados muitos arquivos, em 79,66% das vezes são rejeitados, quando há modificações em alguns arquivos são fechados em 87,27% das vezes. Sendo assim, torna-se interessante que o desenvolvedor, ao submeter seu projeto modifique apenas um arquivo, tendo em vista que as chances de rejeição diminuem em 7% comparado ao *lift* apenas do país (Tabela 17).

**Tabela 17** - Regras de associação relevantes para rejeição quando houverem modificações de arquivos em Reino Unido

#	Antecedente	Consequente	Sup.	Conf.	Lift
1	*United Kingdom* <sup>1</sup> File	Closed	0,77%	69,05%	2,0794
2	*United Kingdom* <sup>Many</sup> Files	Closed	0,62%	79,66%	2,3990
3	*United Kingdom* <sup>Some</sup> Files	closed	0,64%	87,27%	2,6283

Fonte: Elaboração Própria

Quando um desenvolvedor submete seus *commits* no país da República Checa observa-se os valores em que os arquivos são fechados (Tabela 18).

**Tabela 18** - Regras de associação relevantes para aceitação quando houverem modificações de arquivos em República Checa

#	Antecedente	Consequente	Sup.	Conf.	Lift
1	*Czech Republic* <sup>1</sup> File	Closed	0,40%	57,69%	1,7374
2	*Czech Republic* <sup>Many</sup> Files	Closed	0,50%	86,36%	2,6009
3	*Czech Republic* <sup>Some</sup> Files	Closed	0,61%	92,00%	2,7706

Fonte: Elaboração Própria

Na Tabela 18, pode-se ver que quando são modificados muitos arquivos, em 86,36% das vezes são fechados, além disso as chances de rejeição aumentam em 160% comparado ao *lift* do *Status* do país inicialmente (*lift* = 2,2922). Quando um país modifica alguns arquivos, observa-se que as chances de rejeição aumentam em 177%, além de que quando são modificados entre 2 a 4 arquivos são fechados em 92% das vezes. Ao modificar apenas um arquivo este valor diminui em 73%, isso implica que, quando um desenvolvedor for submeter algum projeto, recomenda-se que ele modifique apenas um arquivo, tendo em vista que diminuem as chances de rejeição comparado ao país.

A próxima característica que foi explorada neste estudo de caso, consistiu no total de linhas que o desenvolvedor manipula antes de tentar integrar ao projeto, foram filtradas cerca de 142 regras, e as principais características consideradas foram:

“*Some Lines*”; 2 a 20 linhas e “*Many Lines*” quando o número de linhas manipuladas pelo desenvolvedor forem maiores que 20.

Quando o país de submissão do *Pull Request* for Singapura (*lift* =1,3741), considerando que o *lift* da regra sozinha aumenta em 37% das chances de serem aceitas, quando atenuadas a manipulação de linhas, observa-se que o desenvolvedor manipular algumas linhas aumenta as chances de serem aceitas em 42%. Em contrapartida, ao manipular muitas linhas as chances diminuem em 33% conforme a Tabela 19.

**Tabela 19** – Regras de associação relevantes para aceitação quando houverem modificações de linhas em Singapura.

#	Antecedente	Consequente	Sup.	Conf.	Lift
1	*Singapore* ^ <i>Many Lines</i>	<i>Merged</i>	0,68%	89,47%	1,3395
2	*Singapore* ^ <i>Some Lines</i>	<i>Merged</i>	1,06%	95,24%	1,4258

Fonte: Elaboração Própria

Em Bielorrússia (*lift* =1,3595), as chances de serem aceitas atenua-se apenas quando o desenvolvedor manipula muitas linhas cerca de 37%, a confiança desta regra é de 91,89%. Quando um desenvolvedor manipula entre 2 a 20 linhas, são aceitos em 86,58% das vezes, porém a agregação desta regra não se torna um diferencial muito grande para que sejam aceitas tendo em vista que diminuem as chances de aceitação comparadas ao *lift* apenas do país (Tabela 20).

**Tabela 20** - Regras de associação relevantes para aceitação quando houverem modificações de linhas em Belarus.

#	Antecedente	Consequente	Sup.	Conf.	Lift
1	* Belarus*^ <i>Many Lines</i>	<i>Merged</i>	0,90%	91,89%	1,3757
2	* Belarus*^ <i>Some Lines</i>	<i>Merged</i>	1,14%	89,58%	1,3412

Fonte: Elaboração Própria

Quando o desenvolvedor da Itália manipula muitas linhas (Tabela 21), são rejeitadas em 84,70% das vezes, além disso, considerando que sozinha a Itália tem

*lift* 2,3515, ao atenuar esta regra, o desenvolvedor aumenta em 155% de chances de serem rejeitadas ao manipularem muitas linhas.

Ao manipular algumas linhas, o desenvolvedor oriundo da Itália tem 65,04% de chances de serem rejeitados, além disso suas chances são reduzidas em 95% maior considerando o *lift* do país da Itália sozinho.

**Tabela 21** – Regras de associação relevantes para aceitação quando houverem modificações de linhas na Itália.

#	Antecedente	Consequente	Sup.	Conf.	Lift
1	*Italy*^ <i>Many Lines</i>	<i>Closed</i>	2,05%	84,70%	2,5508
2	*Italy*^ <i>Some Lines</i>	<i>Closed</i>	1,06%	65,04%	1,9587

Fonte: Elaboração Própria

No Reino Unido (Tabela 22), o desenvolvedor quando manipula muitas linhas são rejeitadas em 79,21% das vezes. Além disso, considerando o *lift* apenas do país que tem 132% maior chance de serem rejeitas, ao considerar esta regra as chances atenuam-se em 138%.

Ao manipular algumas linhas, os desenvolvedores diminuem as chances de rejeição em 123% considerando o *lift* apenas do país. Além disso, quando são manipuladas algumas linhas são rejeitadas em 74,19% das vezes.

**Tabela 22** - Regras de associação relevantes para aceitação quando houverem modificações de linhas no Reino Unido

#	Antecedente	Consequente	Sup.	Conf.	Lift
1	*United Kingdom*^ <i>Many Lines</i>	<i>Closed</i>	1,06%	79,21%	2,3854
2	*United Kingdom*^ <i>Some Lines</i>	<i>Closed</i>	0,91%	74,19%	2,2344

Fonte: Elaboração Própria

Na República Checa (*lift* = 2,2922) conforme mostrada na Tabela 23, as características que podem atenuar a rejeição dos projetos consistem nas manipulações de muitas linhas 167% e algumas linhas 189%. Além disso, a confiança gerada na regra para serem rejeitadas consistem em 89,89% das vezes quando forem muitas linhas, e 96% das vezes quando forem manipuladas algumas linhas.

**Tabela 23** - Regras de associação relevantes para aceitação quando houverem modificações de linhas na República Checa

#	Antecedente	Consequente	Sup.	Conf.	Lift
1	*Czech Republic* <sup>^</sup> <i>Many Lines</i>	<i>Closed</i>	0,11%	88,89%	2,6769
2	*Czech Republic* <sup>^</sup> <i>Some Lines</i>	<i>Closed</i>	0,32%	96,00%	2,8911

Fonte: Elaboração Própria

Portanto, considerando os principais atributos para a geração das regras que comprovem quais as características principais atenuam na aceitação e rejeição dos projetos, observam-se que estas características corroboram com a tese de Soares (2017). Como por exemplo a correlação entre *lifetime* que aumentam as chances de solicitação de *Pull Requests* para integração do projeto. Além de que a combinação de características, individualmente importantes, podem aumentar as chances de aceitação do projeto.

Outro tipo de regra gerada consistiu em: *Country*  $\wedge$  *Factors*  $\rightarrow$  *lifetimeMinutes*, na qual os fatores analisados, consistiram em número de arquivos, número de linhas modificadas e linguagem de programação que foram submetidas.

A Tabela 24, tem como objetivo demonstrar as regras que são relevantes em países nos quais existem maiores chances do *lifetime* da regra ocorrer.

**Tabela 24** - Regras de associação relevantes para *lifetime* quando houverem modificações de 1 arquivo em Singapura

#	Antecedente	Consequente	Sup.	Conf.	Lift
1	*Singapore* <sup>^</sup> <i>1 File</i>	<i>very Short</i>	0,93%	83,33%	1,4436
2	*Singapore* <sup>^</sup> <i>Some Files</i>	<i>Very Lengthy</i>	0,13%	21,74%	1,2364
3	*Singapore* <sup>^</sup> <i>Many Files</i>	<i>very Short</i>	0,29%	47,83%	0,8285

Fonte: Elaboração Própria

Quando o país de origem do *Pull Request* for em Singapura e forem modificados, um arquivo, terão um *lifetime* de vida muito mais curto. Considerando apenas as características do *lifetime* nesse país (*lift* = 1,1509), ao agregar a esta regra,

há um aumento na variação do *lift* indicando que esta regra possui uma influência no tempo de vida entre o envio e recebimento de solicitação em períodos bem curtos. Ao serem modificadas entre 2 a 4 linhas, são geradas dependências positivas na regra indicando aumento na taxa da agregação da regra comparada, apenas ao *lifetime* do país. Ou seja, existem 23% maiores chances de quando forem modificadas algumas linhas o tempo de vida tende a ser muito demorado. Em contrapartida, quando o usuário modificar muitas linhas neste país em tempo de vida muito rápido, existem 17,15% menores chances de que esta regra aconteça, comparada apenas ao *lift* do país.

Em Bielorrússia (*lift* = 1,4131), quando o *Pull Request* contiver 1 arquivo modificado, o *lifetime* da regra ocorre em tempo muito curto em cerca de 48% maiores chances. Além disso comparada com o *lifetime* deste país, há um aumento nesta regra indicando ao desenvolvedor, que antes de realizar o *commit* seja mais viável modificar apenas um arquivo.

Comparadas as modificações de alguns arquivos, esta regra acaba tornando-se um pouco mais inviável quando forem analisadas em tempo de vida muito curto, pois, diminuem as chances de aceitação comparadas a regra apenas ao *lifetime* do país. O mesmo vale quando o desenvolvedor modificar muitos arquivos em tempo de vida muito curto, a tabela indica que se torna inviável que neste país sejam modificados muitos arquivos comparadas ao *lifetime* da regra (Tabela 25) .

**Tabela 25** - Regras de associação relevantes para *lifetime* quando houverem modificações de 1 arquivo em Belarus

#	Antecedente	Consequente	Sup.	Conf.	<i>Lift</i>
1	* Belarus <sup>^</sup> 1 File	<i>very Short</i>	0,95%	85,71%	1,4849
2	* Belarus <sup>^</sup> Some Files	<i>very Short</i>	0,71%	79,41%	1,3757
3	* Belarus <sup>^</sup> Many Files	<i>very Short</i>	0,21%	72,73%	1,2599

Fonte: Elaboração Própria

O próximo valor a ser considerado consistiu na quantidade média de arquivos que são manipuladas por desenvolvedores provenientes da Itália (Tabela 26), em tempos de vida: médio, longo, curto, muito curto e muito longo, caracterizados no Quadro 1 da Seção 3.1.

**Tabela 26** - Regras de associação relevantes para *lifetime* quando houverem modificações de 1 arquivo na Itália

#	Antecedente	Consequente	Sup.	Conf.	Lift
1	*Italy*^1 File	Short	0,20%	14,85%	1,1721
2	*Italy*^1 File	Very Short	0,90%	67,33%	1,1663
3	*Italy*^1 File	Medium	0,13%	9,90%	1,1870

Fonte: Elaboração Própria

Pode-se observar nesta regra que quando o desenvolvedor for oriundo da Itália (*lift* = 1,9854) e manipular os arquivos, o *lifetime* desta regra diminui indicando influência quando são modificados um arquivo.

Em contrapartida, quando o desenvolvedor modifica muitos arquivos, o tempo de vida entre a solicitação e recebimento de *Pull Requests* variam possuindo dependências positivas (*lift*>1), e quando atenuadas as regras, indicam que possuem influências quando o *lifetime* = “Medium” e *lifetime* = “lengthy” com variações entre 165 e 116% respectivamente (Tabela 27).

Quando o *lifetime* da regra for em tempos curtos e muito longo, esse fator tende a reduzir o impacto de rejeição em 46% quando *lifetime* = “Short”, e *lifetime* = “Very Lengthy”.

**Tabela 27** - Regras de associação relevantes para *lifetime* quando houverem modificações de muitos arquivos na Itália

#	Antecedente	Consequente	Sup.	Conf.	Lift
1	*Italy*^Many Files	Medium	0,33%	22,12%	2,6524
2	*Italy*^ Many Files	Lengthy	0,12%	7,96%	2,1639
3	*Italy*^ Many Files	Short	0,28%	18,58%	1,4667
4	*Italy*^ Many Files	Very Lengthy	0,25%	16,81%	0,9563
5	*Italy*^ Many Files	very Short	0,52%	34,51%	0,5979

Fonte: Elaboração Própria

Ao considerar que o desenvolvedor da Itália modifique apenas alguns arquivos, observa-se que comparado apenas ao *lift* do *lifetime* (medium = 1,9854, short = 1,4076, very Short=0,8827), quando agregado a regra há um aumento significativo



quando os desenvolvedores da Itália modificam alguns arquivos em tempo de vida, médio, rápido e muito rápido conforme mostra a Tabela 28.

**Tabela 28** - Regras de associação relevantes para *lifetime* quando houverem modificações de alguns arquivos na Itália.

#	Antecedente	Consequente	Sup.	Conf.	Lift
1	*Italy* <sup>^</sup> <i>Some Files</i>	<i>Medium</i>	0,23%	17,00%	2,0381
2	*Italy* <sup>^</sup> <i>Some Files</i>	<i>Short</i>	0,26%	20,00%	1,5785
3	*Italy* <sup>^</sup> <i>Some Files</i>	<i>Very Short</i>	0,70%	53,00%	0,9181

Fonte: Elaboração Própria

Quando os desenvolvedores são oriundos da República Checa, e modificam 1 arquivo observa-se que as variações nos quais são mostradas na Tabela 29 aumentam indicando ao desenvolvedor que o fator *lifetime* tem influência em até 107% quando for em tempo médio por exemplo. Ao analisar outras características das modificações de arquivos, observa-se que que existem aumento quando são realizadas alterações em alguns e até muitos arquivos considerando o fator *lifetime* da regra. Porém quando são modificados muitos arquivos considerando o *lifetime* = “very short”, esse fator diminui em 56,69%. Indicando assim, que o quando o desenvolvedor modifica muitos arquivos, torna-se inviável que o tempo de vida do arquivo seja muito rápido.

**Tabela 29** - Regras de associação relevantes para *lifetime* quando houverem modificações de arquivos na República Checa.

#	Antecedente	Consequente	Sup.	Conf.	Lift
1	*Czech Republic* <sup>^</sup> <i>1 File</i>	<i>Medium</i>	0,12%	17,31%	2,0750
2	*Czech Republic* <sup>^</sup> <i>1 File</i>	<i>Short</i>	0,12%	17,31%	1,3660
3	*Czech Republic* <sup>^</sup> <i>1 File</i>	<i>Very Short</i>	0,40%	57,69%	0,9994
4	*Czech Republic* <sup>^</sup> <i>Many Files</i>	<i>Very Lengthy</i>	0,20%	34,09%	1,9389
5	*Czech Republic* <sup>^</sup> <i>Many Files</i>	<i>Very Short</i>	0,15%	25,00%	0,4331
6	*Czech Republic* <sup>^</sup> <i>Some Files</i>	<i>Medium</i>	0,16%	24,00%	2,8773
7	*Czech Republic* <sup>^</sup> <i>Some Files</i>	<i>Very Lengthy</i>	0,20%	30,00%	1,7062

Fonte: Elaboração Própria

No Reino Unido, considerando que este país possui maiores índices de rejeição conforme abordado inicialmente na Seção 3.1, ao ser aplicado na regra, verificou-se uma diminuição nos índices de rejeição quando são modificados muitos arquivos com *lifetime* da regra “short” e “very short”, em até 60% conforme a Tabela 30. Quando os desenvolvedores modificam alguns arquivos, observa-se que houve um aumento, potencializando as chances de rejeição em até 23% considerando quando *lifetime* = “short”. Ao modificarem apenas um arquivo observou-se que houve um aumento em 8% quando o *lifetime* da regra for muito rápido, e diminui em 32% quando o tempo de vida for muito longo, comparado ao *lift* apenas do *lifetime* do país.

**Tabela 30** - Regras de associação relevantes para *lifetime* quando houverem modificações de arquivos no Reino Unido.

#	Antecedente	Consequente	Sup.	Conf.	Lift
1	*United Kingdom*^1 File	Very Short	0,71%	64,29%	1,1136
2	*United Kingdom*^ 1 File	Very Lengthy	0,19%	16,67%	0,9479
3	*United Kingdom*^ Many Files	Short	0,16%	20,34%	1,6052
4	*United Kingdom*^ Many Files	Very Short	0,42%	54,24%	0,9396
5	*United Kingdom*^ Some Files	Short	0,15%	20,00%	1,5785
6	*United Kingdom*^ Some Files	Very Short	0,50%	69,09%	1,1969

Fonte: Elaboração Própria

O próximo fator analisado em fatores consistiu em número de linhas em que os desenvolvedores dos países manipulam, considerando os status de “*Many Lines*” quando são modificadas mais de 20 linhas e “*Some Lines*” onde o número de linhas modificadas está entre 2 e 20.

Na Tabela 31, ao considerar que em Singapura são manipuladas muitas linhas conforme o antecedente da regra, o *lifetime* aumenta em 66% quando ocorrer entre 1 a 3 dias, e cerca de 39% quando o tempo de vida considerado for maior que 10 dias. Quando o desenvolvedor manipular algumas linhas, as chances aumentam em cerca de 42% em até 24 horas.

**Tabela 31** - Regras de associação relevantes para *lifetime* quando houverem modificações de linhas em Singapura.

#	Antecedente	Consequente	Sup.	Conf.	Lift
1	*Singapore* <sup>^</sup> Many Lines	Short	0,16%	21,05%	1,6616
2	*Singapore* <sup>^</sup> Many Lines	Very Lengthy	0,19%	24,56%	1,3969
3	*Singapore* <sup>^</sup> Some Lines	Very Short	0,91%	82,14%	1,4230

Fonte: Elaboração Própria

Em Bielorrússia (*lift*= 1,3147), quando as pessoas manipulam algumas linhas, possuem 42% maiores chances do seu *lifetime* da regra ser entre 1 a 3 dias comparado ao *lift* apenas do *lifetime*. Considerando que o desenvolvedor manipule muitas linhas, essa chance comparada ao *lift* do *lifetime* diminuiu em 38% conforme a Tabela 32.

**Tabela 32** - Regras de associação relevantes para *lifetime* quando houverem modificações de linhas em Belarus.

#	Antecedente	Consequente	Sup.	Conf.	Lift
1	* Belarus* <sup>^</sup> Some Lines	Very Short	1,05	82,29	1,4256
2	* Belarus* <sup>^</sup> Many Lines	Very Short	0,78	79,73	1,3812

Fonte: Elaboração Própria

Considerando países que possuem altos índices de rejeição, conforme a Tabela 33, observa-se que quando um desenvolvedor for da República Checa e modificar algumas linhas possui 91% maiores chances do *lifetime* ser analisado em tempo médio. Quando o desenvolvedor for do Reino Unido e modificar muitas linhas há 64% maiores chances do *lifetime* ser rápido. Ao submeter *Pull Request* na Itália e forem modificadas muitas linhas há 142% maiores chances do *lifetime* = “medium”. E ao modificar algumas linhas há um aumento em 14% no tempo de vida ser muito rápido.

**Tabela 33** - Regras de associação relevantes para *lifetime* quando houverem modificações de linhas em países com índices de rejeição.

#	Antecedente	Consequente	Sup.	Conf.	Lift
1	*Czech Republic* ^ <i>Some Lines</i>	<i>Medium</i>	0,11%	16,00%	1,9182
2	*United Kingdom* ^ <i>Many Lines</i>	<i>Short</i>	0,28%	20,79%	1,6410
3	*Italy* ^ <i>Many Lines</i>	<i>Medium</i>	0,49%	20,22%	2,4240
4	*Italy* ^ <i>Many Lines</i>	<i>very Short</i>	1,07%	65,85%	1,1408

Fonte: Elaboração Própria

A próxima regra analisada consistiu em *Country*  $\wedge$  *Language*  $\rightarrow$  *Status*, na qual, considerando os países com maiores chances de aceitação indiquem quais são as linguagens de programação que eles mais submetem para integrações ao repositório principal do projeto.

Ao observar a Tabela 34, quando o desenvolvedor de Singapura submete em C#, são aceitos em 100% das vezes, além disso, considerando o *lift* do país (*lift* = 1,3741), ao *lift* desta linguagem (*lift* = 1,3465), quando há integração destes fatores às regras existe um aumento nas chances de aceitação, indicando uma vantagem ao submeter nesta linguagem.

O desenvolvedor de Bielorrússia, ao submeter um *Pull Request* em Ruby, 92,59% das vezes são aceitos, além disso, comparando com o *lift* deste país (*lift* = 1,3595), com o *lift* da linguagem (*lift* = 1,2013), há um aumento considerável da regra (cerca de 38%) indicando um fator em potencial de solicitação neste país.

Quando o desenvolvedor oriundo da Letônia submete um *Pull Request*, observa-se um decréscimo da regra, comparado com os fatores de influência dos atributos individualmente, tornando-se um fator fraco, indicando ao desenvolvedor que não se torna interessante submeter projetos na linguagem Ruby.

**Tabela 34** - *Country ^ Language → Status = "Merged"*

#	Antecedente	Consequente	Sup.	Conf.	Lift
1	*Singapore* ^ C#	<i>Merged</i>	0,16%	100,00%	1,4971
2	*Singapore* ^ Ruby	<i>Merged</i>	1,60%	93,80%	1,4043
3	*Belarus* ^ Ruby	<i>Merged</i>	1,99%	92,59%	1,3862
4	*Latvia* ^ Ruby	<i>Merged</i>	0,11%	88,89%	0,11%

Fonte: Elaboração Própria

Em países que possuem maiores chances de rejeição, observa-se na Tabela 35, as principais linguagens que cada país submete indicando as variações de confiança e *lift* que cada regra possui.

**Tabela 35** - *Country ^ Language → Status = "Closed"*

#	Antecedente	Consequente	Sup.	Conf.	Lift
1	*Italy* ^ Java	<i>Closed</i>	2,95%	88,49%	2,665
2	*United Kingdom* ^ Python	<i>Closed</i>	0,11%	100,00%	3,0116
3	*United Kingdom* ^ PHP	<i>Closed</i>	0,16%	92,31%	2,7799
4	*United Kingdom* ^ Java	<i>Closed</i>	1,63%	82,00%	2,4695
5	*Czech Republic* ^ Java	<i>Closed</i>	1,32%	96,15%	2,8957

Fonte: Elaboração Própria

Ao analisar esta regra, considerando a Itália quando desenvolve na Linguagem Java, são fechadas em 88,49% das vezes, além disso, ao analisar o *lift* das regras individualmente, percebe-se que os valores aumentam, indicando forte influência do status desta regra.

No Reino Unido quando o desenvolvedor submete seus projetos nas linguagens Python, PHP ou Java aumentam as chances de serem fechadas em até 201% (considere a Tabela 35), tornando-se inviável desenvolver nestas linguagens tendo em vista suas altas chances de rejeição.

Quando o desenvolvedor da República Checa submete seu *pull* na linguagem Java, observa-se que em 96,15% das vezes são fechadas, além disso ao realizar a junção destas regras aumenta as chances de rejeição em até 189%.

## 4 CONSIDERAÇÕES FINAIS E RECOMENDAÇÕES

Neste capítulo, serão apresentadas as considerações finais concernentes a este trabalho e recomendações de trabalhos futuros.

### 4.1 CONSIDERAÇÕES FINAIS

O objetivo deste trabalho consistiu em analisar a influência da localização geográfica do *requester* nos cenários de aceitação e tempo de vida do *Pull Request*.

Para cumprir este estudo, foram realizadas análises na base de dados, para compreensão de fatores que podem contribuir na aceitação ou rejeição dos projetos de desenvolvimento considerando o país de origem dos *Pull Requests*. Antes de utilizar a base para a realização destes experimentos, os atributos foram discretizados, organizados na base de dados, para que fossem realizados os estudos para encontrar os valores gerados pelas regras.

Por meio dos estudos que foram realizados na base abordados a partir da Seção 3.1, chega-se à conclusão de que o país possui influência na aceitação e tempo de vida de *Pull Request*, mostrando que existem países com *lifetimes* mais longos assim como existem países com *lifetimes* mais curtos, ao ser comprovada essa hipótese, a Seção 3.2 justificou quais eram os fatores que tornavam as regras mais fortes para aceitação e rejeição, caracterizando elas. Na Seção 3.3 foi realizada a

correlação das regras analisando quais eram os fatores que atenuariam ou potencializariam as chances de aceitação de um projeto.

Como limitações identificadas durante o desenvolvimento deste trabalho, destaca-se inicialmente a organização dos atributos, considerando que quase todos eram numéricos, ajustá-los a um padrão, demandou bastante tempo.

## **4.2 RECOMENDAÇÕES**

Durante o desenvolvimento desta pesquisa, foram observados alguns tópicos que podem ser direcionados para pesquisas em trabalhos futuros, dentre os quais podem se destacar:

- a) Análise da localização geográfica restringindo-se as cidades dos países que possuem as maiores chances de aceitação e rejeição dos projetos, podendo assim refinar o estudo de caso para descobrir quais são as principais características que essas cidades possuem e quais fatores podem potencializar para que essas solicitações sejam aceitas;
- b) Avaliação dos padrões que foram encontrados neste trabalho, considerando o aspecto temporal.



## REFERÊNCIAS

FIGUEIREDO, S.; SANTOS, G.; ROCHA, A. R. **Gerência de Configuração em Ambientes de Desenvolvimento de Software Orientados a Organização**. p. 14, 2004.

GIL, Antônio Carlos. **Métodos e técnicas de pesquisa social / Antônio Carlos Gil**. – 6. Ed. - São Paulo: Atlas 2008.

GONÇALVES, E. C. **Regras de Associações e suas Medidas de Interesse Objetivas e Subjetivas**. INFOCOMP Journal of Computer Science, v. 4, n. 1, p. 26–35, 2004.

GOUSIOS, G.; PINZGER, M.; DEURSEN, A. VAN. **An Exploratory Study of the Pull-based Software Development Model**. ICSE 2014, 2014, New York, NY, USA. *Anais...* New York, NY, USA: ACM, 2014. p. 345–355. Disponível em: <<http://doi.acm.org/10.1145/2568225.2568260>>. Acesso em: 5 jun. 2019.

GITHUB. **About Pull Requests - GitHub Help**. Disponível em: <<https://help.github.com/en/articles/about-pull-requests>>. Acesso em: 12 maio 2019.

MURTA, L. *et al.* **A Caminho da Manutenção de Software Baseado em Componentes via Técnicas de Gerência de Configuração de Software**♣. p. 17, 2004.

PINTO, Gustavo; KAMEI, Fernando K. **Analisando as Contribuições da Comunidade Open Source Brasileira em Projetos Distribuídos de Software Um Estudo Inicial**.

PRIKLADNICKI, R. **MuNDDoS UM MODELO DE REFERÊNCIA PARA DESENVOLVIMENTO DISTRIBUÍDO DE SOFTWARE**. p. 143, 2003.

PRIKLADNICKI, R. *et al.* **Desenvolvimento Distribuído de Software: um Modelo de Classificação dos Níveis de Dispersão dos Stakeholders**. p. 10, 2004.

RASTOGI, A. *et al.* **Relationship Between Geographical Location and Evaluation of Developer Contributions in Github**. ESEM '18, 2018, New York, NY, USA. *Anais...* New York, NY, USA: ACM, 2018. p. 22:1–22:8. Disponível em: <<http://doi.acm.org/10.1145/3239235.3240504>>. Acesso em: 12 maio 2019.

REZENDE, D. A. **Engenharia de Software e Sistemas de Informação**. 3. ed. Rio de Janeiro: Brasport, 2006.

ROMÃO, Wesley et al. **Extração de regras de associação em C&T: O algoritmo Apriori**. XIX Encontro Nacional em Engenharia de Produção, v. 34, p. 37-39, 1999.

SILVA, Edna Lúcia da, MENEZES, Estera Muszkat. **Metodologia da pesquisa e elaboração de dissertação**. 4. ed. Florianópolis: UFSC, 2005.

SOARES, D. M. **On The Nature Of Pull Requests**: A Study About This Collaboration Paradigm Over Open-Source Projects Using Association Rules. Rio de Janeiro: Universidade Federal Fluminense, 2017.

SOARES, D. M. *et al.* **Acceptance factors of Pull Requests in open-source projects**. In: THE 30TH ANNUAL ACM SYMPOSIUM, 2015, Salamanca, Spain. *Anais...* Salamanca, Spain: ACM Press, 2015. p. 1541–1546. Disponível em: <<http://dl.acm.org/citation.cfm?doid=2695664.2695856>>. Acesso em: 5 jun. 2019.

GARG, Anisha Complete guide to Association Rules **Towards Data Science**, 2018. Disponível em: <<https://towardsdatascience.com/association-rules-2-aa9a77241654>> Acesso em: 26 de novembro de 2019.

VASCONCELOS, Livia Maria Rocha de; CARVALHO, Cedric Luiz. **Aplicação de Regras de Associação para Mineração de dados na WEB**. 2004.

WAZLAWICK, R. **Engenharia de Software: Conceitos e práticas**. 1. ed. Rio de Janeiro: Elsevier Brasil, 2013. LOPES, Leandro Teixeira; MAJDENBAUM, Azriel; AUDY, Jorge Luis Nicolas. Uma Proposta para Processo de Requisitos em Ambientes de Desenvolvimento Distribuído de Software. In: **WER**, 2003.

WEKA. Weka 3: Machine Learning Software in Java **waikato**, 2019. Disponível em: <<https://www.cs.waikato.ac.nz/ml/weka/>>. Acesso em: 26 de novembro de 2019.

WEKAPAR. WekaPAR-0.6.5 **GitHub**, 2018. Disponível em: <<https://github.com/gems-uff/wekapar/releases/tag/v0.6.5>> Acesso em: 26 de novembro de 2019.

YU, Yue et al. **Reviewer recommender of pull-requests in GitHub**. In: 2014 IEEE International Conference on Software Maintenance and Evolution. IEEE, 2014. p. 609-612..

## **APÊNDICE**

## **APÊNDICE A– PROJETOS QUE COMPÕEM A BASE DE DADOS**

PROJETOS QUE COMPÕEM A BASE DE DADOS			
#	Projetos	Count	Weight
1	TrinityCore/TrinityCore	283	283.0
2	xbmc/xbmc	678	678.0
3	rails/rails	2561	2561.0
4	zendframework/zf2	173	173.0
5	akka/akka	138	138.0
6	Baystation12/Baystation12	52	52.0
7	diaspora/diaspora	713	713.0
8	adobe/brackets	61	61.0
9	ipython/ipython	90	90.0
10	joomla/joomla-cms	77	77.0
11	mozilla/kuma	63	63.0
12	infinispan/infinispan	1072	1072.0
13	NancyFx/Nancy	169	169.0
14	nodejs/node	83	83.0
15	rapid7/metasploit-framework	5	5.0
16	D-Programming-Language/phobos	147	147.0
17	puppetlabs/puppet	6	6.0
18	ros/rosdistro	404	404.0
19	angular/angular.js	7	7.0
20	scala/scala	29	29.0
21	scikit-learn/scikit-learn	124	124.0
22	appcelerator/titanium_mobile	1	1.0
23	django/django	122	122.0
24	jquery/jquery	44	44.0
25	netty/netty	356	356.0
26	docker/docker	95	95.0

## **APÊNDICE B– LOCALIZAÇÃO GEOGRÁFICA DOS CONTRIBUIDORES**

LOCALIZAÇÃO GEOGRÁFICA DOS CONTRIBUIDORES			
#	Projetos	Count	Weight
1	[*Ukraine*]	190	190.0
2	[*Romania*]	321	321.0
3	[*Slovakia*]	39	39.0
4	[*Poland*]	165	165.0
5	[*Brazil*]	194	194.0
6	[*Spain*]	142	142.0
7	[*China*]	98	98.0
8	[*Belgium*]	163	163.0
9	[*Canada*]	180	180.0
10	[*Germany*]	1163	1163.0
11	[*Czech Republic*]	146	146.0
12	[*Uruguay*]	70	70.0
13	[*Mexico*]	29	29.0
14	[*Portugal*]	12	12.0
15	[*Hungary*]	20	20.0
16	[*Croatia*]	25	25.0
17	[*Russia*]	344	344.0
18	[*Indonesia*]	5	5.0
19	[*United States*]	43	43.0
20	[*Slovenia*]	44	44.0
21	[*Bulgaria*]	67	67.0
22	[*Denmark*]	63	63.0
23	[*Italy*]	314	314.0
24	[*Finland*]	70	70.0
25	[*Norway*]	106	106.0
26	[*Tunisia*]	4	4.0
27	[*Lithuania*]	24	24.0
28	[*France*]	416	416.0
29	[*Cuba*]	3	3.0
30	[*United Kingdom*]	198	198.0
31	[*Austria*]	148	148.0
32	[*New Zealand*]	89	89.0
33	[*Netherlands*]	106	106.0
34	[*Greece*]	40	40.0
35	[*Taiwan*]	75	75.0
36	[*Australia*]	276	276.0
37	[*Switzerland*]	483	483.0
38	[*Iceland*]	2	2.0
39	[*Sweden*]	163	163.0



40	[*Argentina*]	65	65.0
41	[*Lebanon*]	1	1.0
42	[*Israel*]	28	28.0
43	[*Cyprus*]	5	5.0
44	[*Latvia*]	10	10.0
45	[*Serbia*]	8	8.0
46	[*South Korea*]	21	21.0
47	[*Japan*]	385	385.0
48	[*Colombia*]	168	168.0
49	[*Turkey*]	13	13.0
50	[*India*]	258	258.0
51	[*Singapore*]	146	146.0
52	[*Bermuda*]	2	2.0
53	[*Ecuador*]	1	1.0
54	[*South Africa*]	6	6.0
55	[*Belarus*]	174	174.0
56	[*Estonia*]	5	5.0
57	[*Thailand*]	13	13.0
58	[*Georgia*]	1	1.0
59	[*Hong Kong*]	8	8.0
60	[*Ireland*]	31	31.0
61	[*Moldova*]	50	50.0
62	[*Azerbaijan*]	6	6.0
63	[*Netherlands Antilles*]	1	1.0
64	[*Mongolia*]	1	1.0
65	[*Kazakhstan*]	4	4.0
66	[*Peru*]	2	2.0
67	[*Kyrgyzstan*]	1	1.0
68	[*Philippines*]	6	6.0
69	[*Luxembourg*]	6	6.0
70	[*Malaysia*]	1	1.0
71	[*Maldives*]	1	1.0
72	[*Jamaica*]	5	5.0
73	[*Morocco*]	26	26.0
74	[*Egypt*]	4	4.0
75	[*Chile*]	4	4.0
76	[*Iran*]	5	5.0
77	[*Bangladesh*]	4	4.0
78	[*Nepal*]	1	1.0
79	[*Venezuela*]	6	6.0
79	[*Venezuela*]	6	6.0
81	[*Pakistan*]	4	4.0
82	[*Costa Rica*]	1	1.0
83	[*Sri Lanka*]	10	10.0

84	[*Honduras*]	2	2.0
85	[*Guatemala*]	1	1.0
86	[*Paraguay*]	1	1.0
87	[*Macedonia*]	2	2.0
88	[*Kenya*]	1	1.0