

**DEFINIR TÍTULO DO PROJETO @@@**

---

Relatório Científico Final do projeto na modalidade Auxílio à Pesquisa Regular, fomentado pela Fundação de Amparo à Pesquisa do Estado de São Paulo.

---

Pesquisador Responsável: Gustavo Torres Custodio

Santo André, 19 de Janeiro de 2021

# Informações Gerais do Projeto

- Título do projeto:

**Definir título do projeto @@@**

- Nome do pesquisador responsável:

**Gustavo Torres Custodio**

- Instituição sede do projeto:

**Centro de Matemática, Computação e Cognição da Universidade Federal do ABC**

- Equipe de pesquisa:

**Gustavo Torres Custodio**

**Debora Maria Rossi de Medeiros, Dra.**

- Período de vigência:

**01/junho/2015 a 30/maio/2017**

- Período coberto por este relatório científico:

**01/junho/2015 a 30/maio/2017**

## Resumo

Isso é um teste babalsbak das bsak blkab lsab dbaslkb saldbd lsab lasbdlsba ldbas lbds lab ldkasbkldbs alk bdaslkb ldsakb lsdablk bald blkb dlskab lasdbl sablkds ablk b lkasbd lsdbaklb d

## **Abstract**

teste in english

# Sumário

<b>Informações Gerais do Projeto</b>	<b>i</b>
<b>Resumo</b>	<b>ii</b>
<b>Abstract</b>	<b>iii</b>
<b>1 Resumo do projeto proposto</b>	<b>1</b>
<b>2 Apresentação do Problema</b>	<b>2</b>
<b>3 Revisão Bibliográfica Sistemática</b>	<b>3</b>
3.1 Busca em Repositórios . . . . .	3
3.2 Análise de Abstracts . . . . .	4
3.3 Perguntas de Pesquisa . . . . .	4
3.4 Síntese de Resultados . . . . .	5
<b>4 Proposta</b>	<b>7</b>
<b>5 Auxílios Anteriores</b>	<b>8</b>
<b>Referências Bibliográficas</b>	<b>8</b>

# 1 Resumo do projeto proposto

Aqui começa o primeiro capítulo com o resumo do projeto proposto.

## 2 Apresentação do Problema

Aqui começa o terceiro capítulo com a descrição do apoio institucional recebido.

## 3 Revisão Bibliográfica Sistemática

Uma revisão bibliográfica sistemática realiza um estudo na literatura científica sobre um determinado assunto aplicando sistematicamente métodos claramente definidos de busca, seleção e síntese de artigos [1].

Um processo revisão sistemática pode ser sintetizada em 5 passos principais:

1. Definir uma ou mais perguntas de pesquisa.
2. Busca em bases de dados eletrônicas de periódicos científicos usando um conjunto de palavras chave.
3. Inclusão e exclusão de artigos de acordo com critérios de inclusão e exclusão pré-definidos.
4. Avaliação da qualidade metodológica dos artigos e crítica de resultados, descartando artigos que estejam abaixo de um nível de qualidade definido.
5. Apresentação dos resultados, sintetizando os pontos principais dos artigos não descartados em nenhuma das etapas.

No processo de revisão bibliográfica deste trabalho, primeiramente foi realizada uma busca em três diferentes repositórios de periódicos científicos utilizando uma *query* de pesquisa. Alguns dos artigos encontrados foram descartados de acordo com o conteúdo de seus *abstracts*, seguindo um conjunto de critérios de exclusão predefinidos. Os últimos restantes receberam uma pontuação relacionada à qualidade do artigo em relação às perguntas de pesquisa.

### 3.1 BUSCA EM REPOSITÓRIOS

*Science Direct*<sup>1</sup>, ACM<sup>2</sup> e IEEEExplore<sup>3</sup> foram os repositórios escolhidos para a realização da busca por artigos. O *Science Direct* foi escolhido pela sua diversidade de periódicos disponíveis, enquanto o IEEEExplore e a ACM foram escolhidos pelo seu foco em computação e engenharia.

A busca nesses *sites* foi feita utilizando a *query*:

```
(( "evolutionary_algorithms" OR "clustering_algorithms" )  
AND "classifier_ensemble") OR ("evolutionary_algorithms"  
AND ("weak_learner" OR "weak_classifier"))
```

---

<sup>1</sup>Disponível em: <https://www.sciencedirect.com/>

<sup>2</sup>Disponível em: <https://www.acm.org/>

<sup>3</sup>Disponível em: <https://ieeexplore.ieee.org/Xplore/home.jsp>



Na *query* de busca busca-se encontrar a relação de algoritmos evolutivos ou algoritmos de agrupamento com *ensembles* de classificadores. Procura-se também encontrar a relação entre classificadores fracos e algoritmos evolutivos, dado que essa é uma das perguntas de pesquisa escolhidas (Seção 3.3).

A busca retornou um total de 490 resultados. Desses, um total de 478 *papers* restaram após a remoção de artigos repetidos e capítulos de livros encontrados.

## 3.2 ANÁLISE DE ABSTRACTS

Uma avaliação prévia foi realizada nos 478 artigos restantes após a remoção de artigos duplicados e capítulos de livros. Nessa avaliação, os *abstracts* dos artigos restantes são analisados e, se aqueles que apresentam algum dos critérios de exclusão definidos são descartados. Os critérios de exclusão definidos nessa revisão excluem:

- documentos repetidos;
- documentos que não estão em sua versão final;
- artigos que não usam *ensembles* para solução de problemas ou não propõem nenhum novo modelo de *ensemble* de classificadores;
- documentos cujo idioma de escrita não é inglês nem português;
- artigos que abordam os conceitos de forma isolada, em nenhum momento combinando *ensembles* com algoritmos evolutivos ou com algoritmos de agrupamento.

Um total de 119 documentos restaram após a eliminação de artigos que apresentavam um ou mais dos critérios citados. Após essa etapa, foi feita uma leitura completa nos artigos buscando responder um conjunto de perguntas de pesquisa.

## 3.3 PERGUNTAS DE PESQUISA

As perguntas de pesquisa foram definidas buscando entender como otimização por algoritmos evolutivos e agrupamento de dados pode auxiliar na construção de *ensembles* de classificadores.

O processo de revisão bibliográfica trabalhou com as seguintes perguntas:

1. Quais algoritmos evolutivos são combinados com *ensembles*?
2. Quais benefícios de combinar esses algoritmos evolutivos com *ensembles*?
3. Quais técnicas de agrupamento de dados são combinadas com *ensembles*?
4. Quais benefícios de combinar essas técnicas de agrupamento com *ensembles*?
5. Quais critérios são utilizadas para combinar classificadores fracos?

Um critério de pontuação foi definido de acordo com o número das perguntas acima que o trabalho responde, juntamente com os detalhes fornecidos em cada resposta.

O critério de pontuação estabelece um valor máximo de 8 pontos para cada artigo. Aqueles que obtiveram 5 ou mais pontos foram selecionados para a etapa de sintetização dos resultados, sendo que um artigo pode obter no máximo 8 pontos.

As perguntas 1 e 3 contabilizam 1 ponto cada quando suas respostas estão contidas no trabalho analisado, enquanto as perguntas 2, 4, e 5 fornecem 2 pontos no máximo cada. Nas perguntas 1 e 3 só é possível receber 0 ou 1 ponto, o que depende dos algoritmos evolutivos ou técnicas de agrupamento serem nomeados ou não. As demais perguntas podem receber valores intermediários entre 0 e 2 e são pontuadas de acordo com os seguintes critérios:

- 2 As vantagens de utilizar o algoritmo evolutivo no trabalho é apresentada (1 ponto). A função *fitness* do algoritmo evolutivo é descrita (0,5 ponto). O indivíduo/cromossomo que forma uma solução candidata no algoritmo evolutivo é descrito (0,5 ponto).
- 4 As vantagens de utilizar a técnica de agrupamento no trabalho são apresentadas (1 ponto). É descrito como a técnica ajuda a construir o *ensemble de classificadores*.
- 5 Os classificadores fracos que formam o *ensemble* são listados (1 ponto). É descrito como os resultados dos classificadores fracos são combinados (1 ponto).

No total @@@ trabalhos conseguiram 5 pontos ou mais e compuseram a síntese de resultados.

### 3.4 SÍNTESE DE RESULTADOS

A análise completa realizada nos @@@ trabalhos restantes encontrou algoritmos frequentemente utilizados para tarefas específicas na construção de *ensembles*.

A maioria dos algoritmos evolutivos utilizados para otimização de *ensembles* foram Algoritmos Genéticos (AGs) comuns, ou então uma variação deles. Esses trabalhos buscaram utilizar AGs para otimizar tanto a acurácia quanto a diversidade dos *ensembles* de classificadores. Na maioria desses trabalhos, os AGs têm a função de seleção de *features*, isto é, selecionar quais atributos da base de dados serão utilizados para treinamento e teste dos classificadores.

No contexto de seleção de *features*, os cromossomos dos AGs possuem comprimento igual ao número de atributos das instâncias, sendo que cada valor 0 ou 1 do cromossomo indica se o *feature* correspondente foi selecionado para o processo de treinamento e teste dos classificadores base do *ensemble*. A combinação de *features* que resulta no melhor valor de *fitness* é normalmente a selecionada para treinar todos os classificadores, no entanto alguns trabalhos utilizam as  $n$  melhores combinações encontradas pelo AG, sendo que para cada um dos  $n$  classificadores diferentes é utilizada uma das combinações de *features* encontrada.

Muitos trabalhos utilizaram a acurácia média da dos classificadores fracos na etapa de teste, ou então um valor de diversidade entre os classificadores como *fitness*. Aqueles que utilizaram as duas métricas como *fitness*, combinaram elas multiplicando o valor de cada uma por um peso e somando os resultados, outros utilizaram uma abordagem de otimização de múltiplos objetivos, em que a acurácia e a diversidade são duas funções objetivos que devem ser otimizadas simultaneamente. Nesse contexto, o algoritmo mais utilizado foi o *Non-dominated Sorting Genetic Algorithm II* (NSGA-II). Esse algoritmo

@@@

## 4 Proposta

## 5 Auxílios Anteriores

O pesquisador responsável recebeu anteriormente uma bolsa TT-4 da Fapesp para o desenvolvimento de um trabalho junto à Agência de Inovação da UFABC @@@

## Referências Bibliográficas

- [1] SAMPAIO, R. Estudos de revisão sistemática: Um guia para síntese criteriosa da evidência científica. *Revista Brasileira De Fisioterapia - REV BRAS FISIOTER*, v. 11, 02 2007.