XV CONGRESSO INTERNACIONAL DE TECNOLOGIA NA EDUCAÇÃO
EDUCAÇÃO E TECNOLOGIA EM TEMPOS DE MUDANÇA
20 a 22 de setembro de 2017
Centro de Convenções de Pernambuco

MINERAÇÃO DE DADOS ABERTOS DA EDUCAÇÃO BÁSICA E ANÁLISE COM MAPAS DE CALOR

OPEN DATA MINING OF BASIC EDUCATION AND ANALYSIS WITH HEAT MAPS

RAFAELLA NASCIMENTO, GERALDO GOMES DA CRUZ JUNIOR, ROBERTA MACEDO M. GOUVEIA

Resumo

A disponibilização de dados educacionais possibilita, através de métodos de descoberta de conhecimento, uma melhor compreensão e análise da situação educacional em uma determinada região. Este artigo utiliza o processo KDD, com o algoritmo de mineração de dados Naive Bayes, para a identificação de informações extraídas de bases do Censo Escolar de 2012 a 2015 de escolas municipais da cidade do Recife. São criados cenários para análise das instituições de ensino. Com os resultados obtidos são ilustrados mapas de calor, a fim de obter uma visualização geográfica das informações. Alguns cenários estudados, como a presença de acessibilidade, quadras e bibliotecas tiveram pouca evolução durante os 4 anos analisados.
Palavras-chave: Mineração de Dados Educacionais, Educação, Censo Escolar.

Abstract

The provision of educational data enables, through methods of knowledge discovery, a better understanding and analysis of the educational situation in a particular region. This article uses the KDD process with the data mining algorithm Naive Bayes, for identifying information extracted from School Census bases 2012-2015 of municipal schools in the city of Recife. They are created scenarios for analysis of educational institutions. The results obtained are illustrated heat maps in order to obtain a display of geographic information. Some scenarios studied, the presence of accessibility, courts and libraries had little evolution over the four years analyzed.
Key-words: Education Data Mining, Education, School Census.

Introdução

Segundo Dessen e Polonia (2007) as escolas, em união com as famílias, alunos professores e diretores, representam as instituições básicas para a educação e a formação social dos indivíduos. O entendimento dos fatores que constituem estas instituições é essencial para mapear e analisar a qualidade do ensino e recursos oferecidos para melhorar a aprendizagem de uma região.

Uma forma de análise e obtenção de informações significantes relativas à educação, saúde, segurança pública, dentre outros aspectos sociais, que vem se popularizando em todo o mundo é abordagem de dados abertos governamentais. Dados abertos são conteúdos que podem ser livremente acessados, usados, modificados, e compartilhados por qualquer um para

qualquer finalidade, sujeito, no máximo, às exigências que preservam a proveniência (origem dos dados) e abertura (permanecer disponível para a população) [Handbook 2012]. Isto significa que quando há o esforço para a publicação e disseminação das informações do setor público na *web*, permitindo a reutilização e a interação destas, tem-se os Dados Abertos Governamentais [Ribeiro e Almeida 2011]. Dentre esses dados abertos existem os dados educacionais, que possibilitam descobertas de conhecimentos no cenário educacional através de Mineração de Dados Educacionais (MDE) [Nascimento 2016].

A cidade do Recife, capital do estado de Pernambuco, conta com cerca de 317 escolas municipais, 4.957 professores e aproximadamente 88.786 alunos. Estas informações foram obtidas a partir de uma análise simples de bases de dados abertos educacionais disponibilizadas pela prefeitura da cidade, referentes ao Censo Escolar de 2015. Estas bases são formadas por diversas variáveis referentes ao desempenho dos alunos, características do corpo docente, infraestrutura das escolas, entre outras. A análise destas bases possibilita o entendimento mais profundo do ecossistema educacional da região, o desenvolvimento de novas metodologias, tomadas de decisão e ações para a melhoria da educação na cidade.

Cidadãos em todo o mundo devem ter acesso às informações de governos e empresas, de modo que a população possa ter ampla transparência e controle social em torno de bens públicos [Palazzi e Tygel 2014]. A transparência é importante dentro da lógica de melhoria de serviços públicos, melhor gestão dos recursos, promoção da inovação, incremento da segurança e bem-estar social [Dutra e Lopes, 2013].

No Brasil, existem problemas em relação aos dados educacionais pela sua falta de completude, qualidade da informação e formatos de dados inadequados para reuso. Estes problemas criam barreiras para o desenvolvimento de novas soluções [Alcantara et al. 2015]. Uma solução interessante para se gerar informações relevantes a partir desses dados, é a aplicação do processo *Knownledge Discovery in Databases* (KDD), ou simplesmente, processo de descoberta de conhecimento em base de dados. De acordo com Frawley et al. (1992), KDD é a "extração de conhecimento previamente desconhecido, implícito e potencialmente útil, a partir de dados".

Uma das etapas do KDD é a Mineração de Dados (MD), que possui como objetivo revelar padrões e produzir regras através de dados, a fim de avaliar os possíveis resultados [Fayyad et al. 1996]. Para a realização dos estudos de caso desse trabalho utiliza-se o algoritmo de classificação do aprendizado de máquina supervisionado *Naive Bayes* [Yoo e Yang 2015].

Este estudo visa a análise e obtenção de informações a partir das bases de dados educacionais da cidade do Recife/PE utilizando a metodologia do KDD. Também é realizado um estudo temporal, utilizando bases de dados de 2012 a 2015, identificando aspectos de melhorias ou retrocessos na qualidade da infraestrutura oferecida pelas escolas, bem como, são ilustrados gráficos e mapas de calor como forma de representação dos resultados obtidos. Os artefatos e resultados gerados a partir deste estudo buscam uma compreensão do ecossistema da educação municipal na cidade do Recife.

Referencial Teórico

O trabalho de Fritsch et al. (2014) tem como objetivo discutir indicadores de qualidade da educação, com ênfase nas taxas de defasagem idade-série no ensino médio, em três escolas públicas estaduais. Para esta análise são utilizados dados abertos educacionais. Conclui-se que as políticas educacionais são insuficientes ou ineficazes, principalmente por não considerarem as diferenças existentes entre as realidades de seus estudantes.

No estudo de Ferreira (2015) o foco é a identificação de fatores relacionados à conclusão do Ensino Fundamental utilizando técnicas de mineração de dados aplicadas aos dados do Censo Escolar da educação básica do INEP de 2014. Foram analisados dados educacionais relativos à estudantes, turmas, escolas e docentes de todo o Brasil. Encontrou-se evidências de que necessidades especiais dos alunos estejam ligadas com a não conclusão do Ensino Fundamental e de que a cor/raça branca, aulas de inglês, espanhol, artes e outras disciplinas não obrigatórias estejam associadas à conclusão do Ensino Fundamental.

ISSN: 1984-6355

O artigo de Silva (2014) utiliza uma tarefa da Mineração de Dados conhecida por Associação de Dados para encontrar padrões de regras nos resultados de provas e questionários socioeconômicos do Exame Nacional de Ensino Médio (ENADE). As diferentes parametrizações do algoritmo de associação *Apriori* permite descobrir problemas na educação nacional, como: (I) alunos que estudam em escola pública, em geral, têm desempenho regular no exame; (II) pais com estudo do primeiro grau têm filhos que estudam em escola pública e com desempenho regular no exame.

No trabalho de Nascimento (2016) é utilizada a mineração de dados da educação básica, envolvendo ensino infantil, fundamental e médio, extraindo informações de bases de dados do Censo Escolar de quatro anos. A autora fez uso tanto do aprendizado de máquina supervisionado, com utilização dos algoritmos de classificação *J48* e o *Naive Bayes,* como também fez uso do aprendizado não supervisionado, com a utilização do algoritmo de associação *Apriori*. Após a obtenção dos resultados, é desenvolvido um sistema de representação geográfica de forma intuitiva, transformando as informações em mapas de calor. Pode-se identificar que as escolas urbanas possuem melhor resultados se comparadas às escolas rurais. Este cenário não apresenta diferenças se as escolas forem analisadas quanto à presença de atividades complementares, como esportes, leitura e inclusão digital, com destaque para as escolas municipais do ambiente rural, que apresentam maior carência neste cenário.

Metodologia

O KDD é uma técnica que possibilita analisar dados para gerar descobertas que podem ajudar na tomada de decisão, otimizando os processos e retornando de forma eficiente a informação para que se possa definir a estratégia mais adequada para se aplicar em determinado cenário. As fases deste processo são as seguintes: seleção dos dados, pré-processamento dos dados, transformação, mineração de dados e análise.

Na seleção, um conjunto de dados sobre o qual se pretende executar o processo é definido, formando um domínio no qual a limpeza deve ser realizada. O âmbito de estudo deste trabalho são as escolas municipais do Recife-PE dos anos de 2012 a 2015. É observado que cada base de dados varia em quantidade de atributos presentes e quantidade de registros (veja Tabela 1).

Tabela 1. Total de registros presentes nas bases de dados.

2012	2013	2014	2015
307 registros	319 registros	330 registros	317 registros
Total: 1273 registros			

No pré-processamento e limpeza, as informações consideradas desnecessárias são removidas e são adotadas estratégias para manusear dados inconsistentes ou faltantes [Refaat 2010]. Também, são utilizados métodos de redução para utilizar apenas atributos relevantes, visando com isto melhorar o desempenho dos algoritmos de análise. Também são aplicadas técnicas de padronização de valores de atributos, por exemplo, um atributo que classifica gênero pode ter valores “Feminino e Masculino” ou “M e H”. Sendo assim, usa-se mecanismos para torná-los correspondentes.

Nesta etapa, é utilizado um *script* para tratar os dados nulos, o qual consiste em substituir as ocorrências vazias nas bases de dados pelo caractere “?”, pois é a forma que a ferramenta de mineração de dados *Weka* interpreta esses valores (Figura 1). Também é feita a exclusão dos atributos que não são relevantes para estudo. Por fim, é selecionado um total de 16 atributos (veja Tabela 2).

Tabela 2. Atributos selecionados após fase de limpeza.

Atributos selecionados		
Ano	Quadra_coberta	Acesso_internet
Agua_rede_publica	Quadra_descoberta	Banda_larga
Esgoto_rede_publica	Dependencia_biblioteca	Laboratorio_informatica
Lixo_coleta_periodica	Dependencia_leitura	Laboratorio_ciencias

ISSN: 1984-6355

Dependencia_adequada_deficiencia	Num_computadores	

A transformação é a fase após o pré-processamento e limpeza. Ela é necessária pois os dados podem não estar em formato adequado para que os algoritmos de mineração os utilizem, sendo necessário a transformação. Como exemplo, alguns algoritmos trabalham com valores numéricos e outros com valores categóricos, sendo necessário transformar os valores numéricos em categóricos ou os categóricos em valores numéricos.

Nesta etapa é utilizado um *script* para transformar os valores contínuos em categóricos, como exemplo, para o atributo "Num_computadores" cria-se categorias, tais são: MaisDe30, Entre21E30, Entre11E20, Ate10 e Nenhum (Figura 1).

```php
ini_set('memory_limit', '-1');
$var = 2;

for ($i = 0; $i < $var; $i++){
  $f="escolas.csv";
  $file=file_get_contents($f);
  $file = str_replace(",,", ",?,", $file);
  file_put_contents($f, $file);
}
```

```php
$read = fopen('escolas.csv', 'r');
$write = fopen('escolas_limpos.csv', 'w');
if ($write && $read) {
    while(($data = fgetcsv($read, 0, ";")) !== FALSE){
      if ($data[18] <> "?"){
        if ($data[18] > 30 ){
          $data[18] = "MaisDe30";
        }
        elseif ($data[18] > 20 && $data[18] <= 30 ){
          $data[18] = "Entre21E30";
        }
        elseif ($data[18] > 10 && $data[18] <= 20 ){
          $data[18] = "Entre11E20";
        }
      }
[...]
      fputcsv($write, $data);
    }
}
fclose($write);
fclose($read);
```

Figura 1. *Script* para tratar valores nulos (A) e trecho do *script* para transformar valores numéricos em categóricos (B).

Após a limpeza e a transformação dos dados, cria-se um arquivo de dados em formato *arff*, do inglês *Attribute Relation File Format*, que é o formato padrão de arquivos da ferramenta *Weka* (Figura 2). Neste arquivo é necessário estabelecer os seguintes cabeçalhos: relação, um nome para o conjunto de dados é definido na primeira linha do arquivo; os atributos são declarados através de uma sequência ordenada de atributos seguida do tipo; conjunto de dados, declarado em uma única linha por *@data*, é onde delimita o início dos dados de instância; dados de instância, são declarados um por linha e deve-se separar os atributos com vírgula. Com o arquivo *arff* criado é então carregado na ferramenta *Weka*, a qual possibilita ter uma visão inicial dos registros existentes, como total de instâncias e atributos, número de registros perdidos e gráficos dos atributos e classes (Figura 2).

```
@relation schools

@attribute ano {2012,2013,2014,2015},
@attribute agua_rede_publica {0,1},
@attribute esgoto_rede_publica {0,1},
@attribute lixo_coleta_periodica {0,1},
@attribute dependencia_laboratorio_informatica {0,1},
@attribute dependencia_laboratorio_ciencias {0,1},
@attribute dependencia_quadra_coberta {0,1},
[...]
@data
2012,1,1,0,1,1,0,1,1,1,0,0,0,0,0,0,11Ate20,Entre11E20
2012,1,0,1,1,1,0,1,0,0,1,0,0,1,1,0,11Ate20,Entre21E30
2012,1,0,1,1,1,0,1,0,0,1,0,0,0,0,0,11Ate20,Entre11E20
2012,1,0,1,1,0,0,1,1,1,0,0,0,0,1,1,6Ate10,MaisDe30,1,
2012,1,0,1,1,1,0,0,0,0,1,1,0,0,1,0,6Ate10,Entre11E20,
2012,1,1,0,1,0,0,0,1,1,0,0,0,0,0,0,6Ate10,MaisDe30,1,
[...]
```

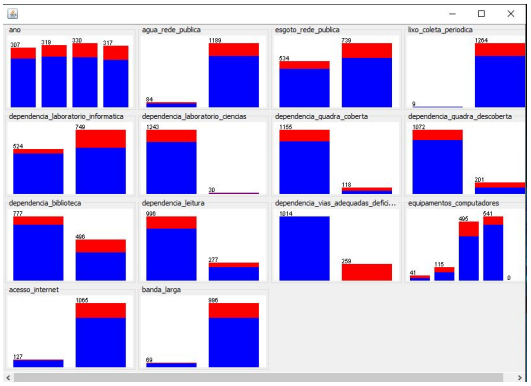


Figura 2. Trechos do arquivo *arff* (A) e gráficos gerados na ferramenta *Weka* com base na classe e atributos definidos na visualização inicial (B).

A mineração de dados (MD) é a principal fase do processo KDD. É uma forma de explorar e analisar bases de dados, buscando extrair conhecimentos como regras, padrões e desvios [Kampff 2009]. A escolha da técnica de mineração mais adequada depende de aspectos como a área do problema e dos dados disponíveis.

ISSN: 1984-6355

Neste trabalho, a MD é executada a partir da ferramenta *Weka* versão 3.7, a qual possui diversos algoritmos de aprendizado de máquina que podem ser utilizados para extrair informações relevantes de uma base de dados. A escolha da ferramenta é justificada por ser *Open Source*, possuir algoritmos de diferentes abordagens da mineração de dados, disponibilidade de recursos estatísticos e análise dos dados, além de uma interface fácil de ser entendida.

Dentre os algoritmos presentes na ferramenta, o *Naive Bayes* é escolhido na fase de mineração desse estudo, visto que apresenta um modelo de resultados numéricos que pode ser facilmente convertido em gráficos e tabelas para fazer análises quantitativas dos principais resultados. Isto facilita também pois é um algoritmo de classificação, e em relação a uma classe são gerados os resultados para os demais atributos.

Apesar de utilizar a ferramenta computacional para executar o algoritmo de mineração, os resultados ainda precisam passar por uma análise do especialista. No entanto, a mineração de dados é significativa no processo de descoberta de conhecimento, permitindo que os esforços se concentrem apenas em partes mais significativas dos dados.

Resultados

Como o objetivo é informar sob o panorama temporal dos dados – de 2012 a 2015, é realizada a execução do *Naive Bayes*, pode-se realizar estudos sobre os resultados obtidos. Estas informações são analisadas sobre os seguintes cenários: Infraestrutura Básica, Ciência e Tecnologia, Leitura e Esportes.

Em relação à infraestrutura básica, são definidos os seguintes atributos para análise: a existência de rede pública de água, de rede de esgoto, de coleta periódica do lixo e da existência de dependências adequadas às pessoas com deficiência (Tabela 3).

Tabela 3. Análise do cenário Infraestrutura Básica.

Infraestrutura básica	2012	2013	2014	2015
Rede pública de água	93,2%	92,5%	93,9%	92,7%
Rede de esgoto	56,6%	55,7%	57,5%	62,06%
Coleta periódica do lixo	99,3%	98,7%	99,06%	98,7%
Dependências acessíveis	18,44%	18,38%	21,68%	23,51%

Como pode-se observar na Tabela 3, a ocorrência de escolas que possuem rede pública de água e coleta periódica do lixo são altas nos 4 anos de análise, e não apresentam mudanças significativas na análise temporal. Porém, destacam-se dois atributos: apesar de representar apenas um pouco mais de 50% durante os anos, a presença de rede de esgoto nas escolas de 2012 a 2015 teve um pequeno aumento (pouco mais de 5%); em relação a presença de dependências acessíveis, percebe-se uma carência nas escolas municipais do Recife, no entanto, houve um pequeno aumento no ano de 2014 e que se segue no ano de 2015, ou seja, em relação a 2012 o aumento foi de pouco mais de 5%.

Em relação ao cenário Ciência e Tecnologia, são definidos os seguintes atributos para análise: a existência de laboratórios de informática e de ciências, números de computadores, presença de internet e de conexão banda larga (Tabela 4).

Tabela 4. Análise do cenário Ciência e Tecnologia.

Ciência e Tecnologia	2012	2013	2014	2015
Laboratório de Ciências	2,91%	1,55%	2,40%	3,76%
Laboratório de Informática	65,69%	63,86%	59,33%	46,39%
Internet	75,52%	91,30%	92,90%	95,73%
Conexão banda larga	93,08%	93,43%	93,07%	93,17%
Número de computadores (Até 10)	32,87%	36,42%	51,43%	58,11%

Tendo em vista os resultados da Tabela 4, nota-se que em relação à presença de laboratórios de ciências, apresar de apresentar baixas ocorrências, as escolas municipais

ISSN: 1984-6355

apresentam um pequeno aumento entre os anos de 2012 e 2015. Já em relação aos laboratórios de informática, é identificado um decréscimo durante os anos (pouco mais de 19%). Quanto a presença de internet nas escolas, houve um aumento de mais de 20% nas ocorrências entre os 4 anos de estudo, assim como a presença de computadores (até 10), que aumentou pouco mais de 25%.

Em relação ao cenário Leitura e Esportes, são definidos os seguintes atributos para análise: a existência de quadras de esportes, de bibliotecas e salas de leitura (Tabela 5).

Tabela 5. Análise do cenário Leitura e Esportes.

Leitura e Esportes	2012	2013	2014	2015
Quadra coberta	9,38%	9,34%	9,93%	9,40%
Quadra descoberta	15,85%	14,01%	17,16%	16,92%
Biblioteca	38,18%	37,07%	39,15%	41,69%
Sala de leitura	36,86%	23,08%	19,27%	18,80%

Como pode-se observar, a ocorrência de quadras, sendo cobertas ou descobertas, são baixas quando analisados os dados, tendo pequeno acréscimo entre o ano de 2012 e 2015. Em relação a presença de biblioteca nas escolas, nota-se uma melhor presença deste atributo, possuindo também pequeno acréscimo entre os 4 anos (pouco mais de 5%). Já em relação a presença de sala de leitura, as escolas municipais apresentam um decréscimo com o passar dos anos: em 2012 com 36,86%; em 2015 com 18,80% (redução de quase metade das ocorrências).

Após os resultados da fase de mineração de dados, busca-se representar graficamente o cenário atual das escolas municipais do Recife com base nos dados fornecidos pelo Censo Escolar de 2015. Para isto, é utilizada a ferramenta *CartoDB* versão 3.0.1, a qual possibilita criar mapas de calor em relação a algum cenário definido. Para plotar os pontos nos mapas é importante que a base contenha dados de geolocalização, como pontos de latitude e longitude (Figura 4). Ao selecionar a opção de exibição em mapas de calor (*Heatmap*), são feitos ajustes nos parâmetros da ferramenta, para que os dados não percam a sua representatividade real.

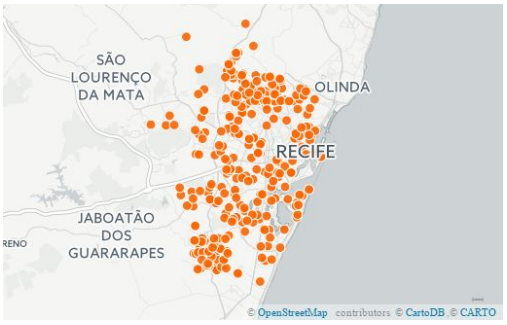
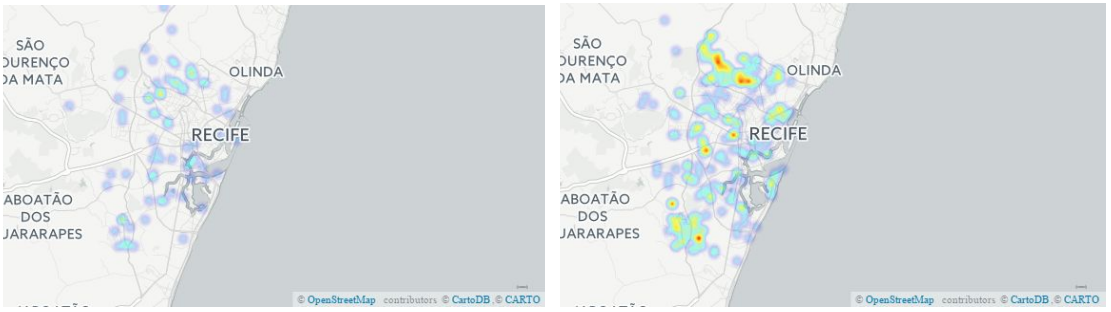


Figura 4. Escolas municipais do Recife representadas por pontos no mapa (2015).

Analisando em relação à existência de dependências adequadas às pessoas com deficiência (Figura 5), exibe que existem mais escolas sem acessibilidade do que com acessibilidade.



ISSN: 1984-6355

Figura 5. Mapa de calor das escolas municipais do Recife que possuem dependências acessíveis (A) e que não possuem dependências acessíveis (B) para o ano de 2015.

Em relação à existência de laboratórios de informática e internet, o cenário apresenta uma melhora significativa se comparada ao cenário da acessibilidade (Figura 6). Verifica-se mais pontos quentes no mapa de calor, ou seja, maior presença de escolas que possuem estes dois atributos.

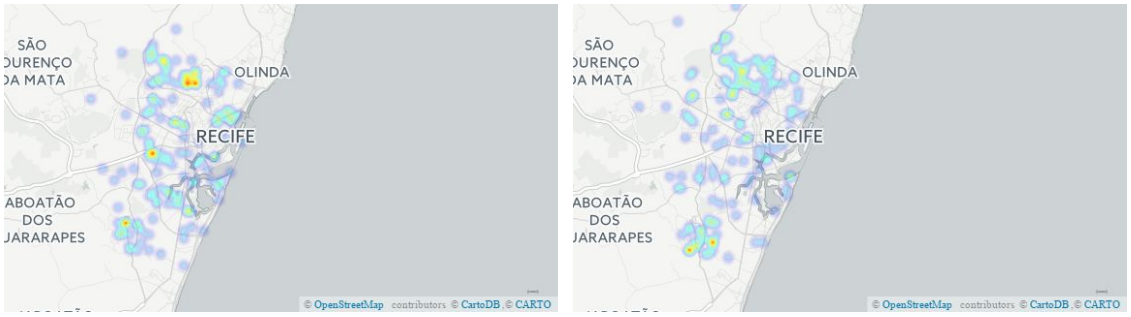


Figura 6. Mapa de calor das escolas municipais do Recife que possuem internet e laboratórios de informática (A) e possuem internet e não possuem laboratórios de informática (B) para o ano de 2015.

Já em relação à existência de quadra de esportes e bibliotecas, o cenário apresenta uma grande necessidade de melhorias, uma vez que a presença destes atributos é pouco representada no mapa de calor (Figura 7).

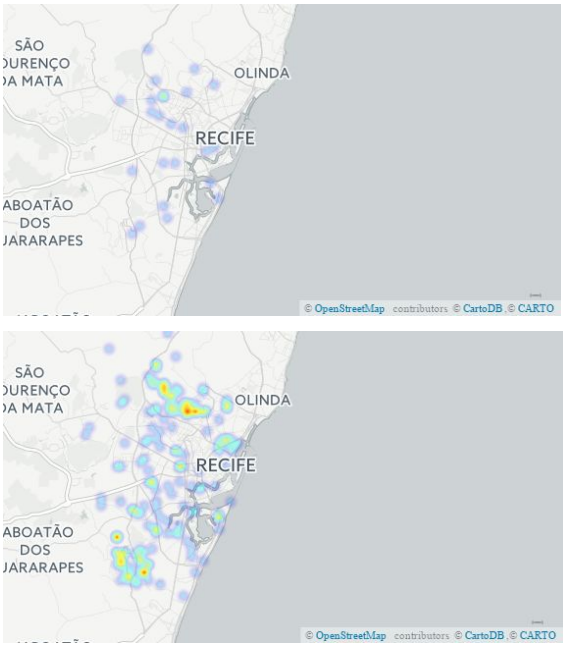


Figura 7. Mapa de calor das escolas municipais do Recife que possuem quadra e biblioteca (A) e que não possuem quadra e biblioteca (B) para o ano de 2015

Considerações Finais

O conhecimento obtido após a mineração de dados e a representação por mapas de calor é bastante relevante, uma vez que informam sobre os três cenários definidos para fins de estudo neste trabalho (Infraestrutura Básica, Ciência e Tecnologia, Leitura e Esportes).

Como principais resultados, foi possível avaliar a infraestrutura básica das escolas, principalmente em relação à acessibilidade, dessa forma, observa-se necessidades de melhorias já que são baixas as ocorrências de escolas com dependências adequadas à pessoas com deficiência. A educação é um direito de todos [Brasil 2014], e isto significa que não deve haver barreias para que qualquer pessoa tenha acesso, sendo assim, trazer melhorias no cenário da acessibilidade promove a inclusão.

Em relação ao incentivo à leitura e esportes o cenário não apresenta diferenças. É muito baixa a presença de escolas que possuem quadras de esportes e bibliotecas ou salas de leitura. Promover estas atividades contribuem para a educação do aluno, que tem contato com atividades que estimularão o seu intelecto. Quanto à presença de internet nas escolas a situação apresenta melhorias quando comparada aos outros cenários, porém a presença de laboratórios ainda é muito baixa para tratar a inclusão digital nessas escolas.

Portanto, percebe-se que as políticas públicas na área da educação para as escolas municipais da cidade do Recife precisam ser revistas, receber mais investimentos e aplicadas. Isto traria um benefício para todos os envolvidos neste âmbito, seja alunos, profissionais da educação, familiares e comunidade.

Referências

Alcantara, W., Bandeira, J., Barbosa, A., Lima, A., Ávila, T., Bittencourt, I., & Isotani, S. (2015). Desafios no uso de Dados Abertos Conectados na Educação Brasileira. In: Anais do DesafiE - 4º Workshop de Desafios da Computação Aplicada à Educação. Congresso da Sociedade Brasileira de Computação.

BRASIL, Lei nº 12.960, de 27 de março de 2014. Estabelece as diretrizes e bases da educação nacional, 2014.

Dessen, M. A., & Polonia, A. C. (2007). A família e a escola como contextos de desenvolvimento humano. Paidéia, 2007, 17(36), 21-32

Dutra, C. C., & Lopes, K. M. G. (2013). Dados abertos: Uma forma inovadora de transparência. VI Congresso CONSAD de Gestão Pública.

Fayyad, U., Piatetsky-Shapiro, G., & Smyth, P. (1996). From data mining to knowledge discovery in databases. AI magazine, 17(3), 37.

Ferreira, G. (2015, October). Investigação acerca dos fatores determinantes para a conclusão do Ensino Fundamental utilizando Mineração de Dados Educacionais no Censo Escolar da Educação Básica do INEP 2014. In: Anais dos Workshops do Congresso Brasileiro de Informática na Educação (Vol. 4, No. 1, p. 1034).

Frawley, W. J., Piatetsky-Shapiro, G., & Matheus, C. J. (1992). Knowledge discovery in databases: An overview. AI magazine, 13(3), 57.

Fritsch, R., Vitelli, R., & Rocha, C. S. (2014). Defasagem idade-série em escolas estaduais de ensino médio do Rio Grande do Sul. Revista Brasileira de Estudos Pedagógicos RBEPINEP, 95, 218-236.

Handbook, O. D. (2012). What is open data. Open Knowledge Foundation.

Kampff, A. J. C. (2009). Mineração de dados educacionais para geração de alertas em ambientes virtuais de aprendizagem como apoio à prática docente.

Nascimento, Rafaella L. S. (2016). Mineração de Dados Educacionais e Visualização de Informações Geográficas Utilizando Mapas de Calor [monografia do curso Sistemas de Informação]. Recife (PE): Universidade Federal Rural de Pernambuco.

Palazzi, D., & Tygel, A. (2014). Visualização de Dados Estatísticos Representados como Dados Abertos Ligados.

Refaat, M. (2010). Data preparation for data mining using SAS. Morgan Kaufmann.

Ribeiro, C. J. S., & Almeida, R. F. (2011). Dados Abertos Governamentais (Open Government Data): Instrumento para Exercício de Cidadania pela Sociedade. Encontro Nacional de Pesquisa em Ciência da Informação, 12, 2568-2580.

Silva, L. A., Morino, A. H., & Sato, T. M. C. (2014). Prática de Mineração de Dados no Exame Nacional do Ensino Médio. In Anais dos Workshops do Congresso Brasileiro de Informática na Educação (Vol. 3, No. 1, p. 651).

Yoo, J. Y., & Yang, D. (2015). Classification Scheme of Unstructured Text Document using TF-IDF and Naive Bayes Classifier. Advanced Science and Technology Letters Vol.111 (COMCOMS 2015), pp.263-266.