

DESENVOLVIMENTO DE UMA PLATAFORMA WEB PARA CIÊNCIA DE DADOS APLICADA À SAÚDE MATERNO INFANTIL

Caio Augusto de Souza Mota¹, Carlos Eduardo Beluzo²,
Lavinia Pedrosa Trabuco³, Adriano Souza⁴, Luciana Alves⁵, Tiago Carvalho⁶

1 Graduando em Tecnologia de Análise e Desenvolvimento de Sistemas, Bolsista PIBIFSP, IFSP, Campus Campinas caiomota802@gmail.com

2 Prof. Orientador do PIBIFSP no IFSP - Campus Campinas beluzo@ifsp.edu.br

3 Graduando em Tecnologia de Análise e Desenvolvimento de Sistemas, IFSP, Campus Campinas laviniaapedrosat@gmail.com

4 Pesquisador Colaborador silva_souza_adriano@gmail.com

5 Profa. colaboradora do NEPO/UNICAMP luciana@nepo.unicamp.br

6 Prof. colaborador do IFSP Campus Campinas tiagoic@ifsp.edu.br

Área de conhecimento (Tabela CNPq): 1.03.03.04-9 Sistemas de Informação

Apresentado no 10º Congresso de Inovação, Ciência e Tecnologia do IFSP
27 e 28 de novembro de 2019- Sorocaba-SP, Brasil

RESUMO: A mortalidade infantil é um problema no mundo todo e sua redução é de interesse mundial. Cerca de 46% das mortes no mundo acontecem até os cinco anos e grande parte desses óbitos está concentrada no primeiro dia e primeira semana de vida. Neste contexto, é importante reconhecer, a necessidade do uso de ferramentas especializadas que possam potencializar o poder dos estudos, de modo a permitir a visualização e manipulação de dados provenientes de grandes segmentos populacionais. Como parte fundamental do projeto de pesquisa *"Plataforma de Apoio à Decisão para Políticas Públicas de Saúde Gestacional Baseada em Técnicas de Visualização de Informações e Aprendizado de Máquina"*, em desenvolvimento no IFSP, está sendo implementada neste trabalho, a infraestrutura lógica para a plataforma web do referido projeto. Desenvolvida utilizando o framework web *Django*, em *Python*, e acoplado o gerenciador de banco de dados *PostgreSQL*, a arquitetura da plataforma está sendo implementada de modo a permitir que futuramente seja possível abrigar métodos de predição de mortalidade e métodos de visualização de dados. A implementação destes métodos está fora do escopo deste trabalho, porém um mecanismo de controle de acesso está sendo implementado, para permitir que usuários possam carregar estes métodos na plataforma.

PALAVRAS-CHAVE: sistema web; visualização de dados; aprendizado de máquina; mortalidade neonatal; demografia.

DEVELOPING A WEB PLATFORM TO DATA SCIENCE APPLIED TO CHILD MATERNAL HEALTH

ABSTRACT: Child mortality is a worldwide problem and its reduction is of worldwide interest. About 46% of deaths worldwide occur until the age of five, most of these deaths are concentrated in the first day and first week of life. In this context, it is important to recognize the need to use specialized tools so that they can leverage the power of studies to allow visualization and manipulation of data from large population segments. As part of the research project *"Decision Support Platform for Public Gestational Health Policy Based on Information Visualization and Machine Learning Techniques"*, under development at IFSP, in this work, the logical infrastructure for the web platform of that project is being implemented. Developed using *Python Django* web framework and coupled with *PostgreSQL*, the platform architecture is being implemented to allow that mortality prediction and data visualization methods, be loaded on the platform as modules (or functionalities). The

implementation of these methods is beyond the scope of this work, but an access control mechanism is being implemented to allow users to load these modules in the platform.

KEYWORDS: web system; data visualization; machine learning; neonatal mortality; demography.

INTRODUÇÃO

A redução neonatal é de extrema importância no mundo. Cerca de 46% das mortes no mundo acontecem entre os menores de cinco anos e grande parte está concentrada nos primeiros dias de vida. Esse período de 28 dias iniciais é o momento mais vulnerável para a sobrevivência da criança (LANSKY et al., 2014). Neste sentido, o uso de técnicas de aprendizado de máquina aplicadas a este contexto é uma proposta inovadora à realidade brasileira, logo é importante reconhecer, a necessidade do uso de ferramentas especializadas para que possam potencializar o poder dos estudos, de modo a permitir a visualização e manipulação de dados provenientes de grandes segmentos populacionais, possibilitando assim a formulação de indicadores de acompanhamento (FRIAS et al., 2017).

Para esse problema está sendo desenvolvido um projeto no IFSP para construir uma *Plataforma de Apoio à Decisão para Políticas Públicas de Saúde Gestacional Baseada em Técnicas de Visualização de Informações e Aprendizado de Máquina*. Dentro deste projeto, o objetivo do presente trabalho é definir e implementar a infraestrutura lógica desta plataforma web. O desenvolvimento dos métodos de predição e das ferramentas de visualização estão fora do escopo deste trabalho, porém a plataforma web deve estar preparada para abrigar estes recursos.

MATERIAL E MÉTODOS

A infraestrutura lógica para a plataforma web está sendo desenvolvida utilizando o framework web *Django* (DJANGO, 2019), em *Python* (PYTHON, 2019), e acoplado o gerenciador de banco de dados *PostgreSQL* (POSTGRESQL, 2019). A plataforma web terá funcionalidades para autenticação de usuário, execução de uma funcionalidade para predição de morte neonatal, e recursos de visualização de dados. Estas duas últimas serão acopladas na plataforma futuramente.

Para isso, a arquitetura da plataforma envolve uma solução *Fullstack* que está sendo implementada com o *framework web Django*, na linguagem de programação *Python*. A arquitetura do projeto baseia-se no padrão *Model-View-Controller*, porém o *Django* se organiza com algumas diferenças, melhor representado como “MTV” (*Model-Template-View*). As “*Views*” por exemplo representam os dados que são apresentados ao usuário. Além disso existe também um diretório somente para os “*Templates*”. O “*Controller*” é o próprio framework, que envia uma requisição para a “*View*” apropriada, de acordo com a configuração de “*URL*” do *Django* (DJANGO, 2019).

Dentre os diversos arquivos da estrutura básica da plataforma, o arquivo “*model.py*” é responsável pela criação e inserção de dados nas tabelas do *PostgreSQL*. Dentro do diretório *website/* existe subdiretório *templates/* que contém o código *HTML* de todas as páginas como a *index.html*, *about.html*, *visualization_list.html*, *login.html*, etc. O subdiretório *static/* é padrão do *Django*, e neste diretório estão localizados todos os arquivos estáticos utilizados pela plataforma, como por exemplos os arquivos *CSS*, *JavaScript*, *JSON* e bibliotecas *Python* (libs).

Os novos módulos de visualizações já podem ser carregados na plataforma em formato *zip*, e ficam armazenadas no diretório *utils/*, e posteriormente são descompactadas. Para armazenamento de dados o gerenciador de banco de dados *PostgreSQL* está sendo utilizado. Ele é integrado ao *Django* por meio das bibliotecas *django_database_url* e *psycopg2* que são responsáveis em interpretar os códigos em *Python* dentro do arquivo “*website/models*” e converter o código para a linguagem do *PostgreSQL* assim criando e inserindo dados nas tabelas. No arquivo “.env” são feitas as configurações de conexão com o banco de dados, configurações como *root* e *base de dados*.

Utilizando *Bootstrap* (BOOTSTRAP, 2019) a fim de se agregar a plataforma elementos e modelos *CSS* que possam ser adaptáveis no decorrer da estilização. O *Bootstrap* possui um sistema de *grade* que permite transformar uma página comum em uma página responsiva, onde poderá ser vista

de uma forma esteticamente agradável tanto em um computador quanto em um celular, a utilização do *Bootstrap* traz uma rentabilidade no processo de estilização tendo um foco na estética para que se possa ter uma maior taxa de satisfação dos usuários.

O modelo de predição será posteriormente implementado e disponibilizadas para serem integradas na plataforma em formato de bibliotecas. As ferramentas de visualização de dados serão desenvolvidas usando *D3JS* (D3JS, 2019) e *ReactJS* (REACTJS, 2019) que são bibliotecas JavaScript utilizados para criar gráficos interativos. Essas bibliotecas consomem um arquivo em formato *JSON*, que é carregado pelo usuário juntamente com a biblioteca e um arquivo HTML.

RESULTADOS E DISCUSSÃO

Resultados esperados é que ao final o usuário, após ter suas credenciais validadas, tenha duas opções, a primeira é executar um modelo de predição (ou visualização) previamente cadastrado, ou realizar o carregamento de uma nova técnica. Até o momento algumas telas da plataforma já estão implementadas. Temos um modelo de visualização de dados (mapa) para visualização de taxa de mortalidade neonatal no Brasil. Já está implementado o mecanismo para controle acesso de usuários, assim como funcionalidade que permite fazer carregamento de modelos de predição.

Para o gerenciamento de versões de documentos e códigos fonte esta sendo o utilizando o GitHub (<https://github.com/cbeluzo/GCE/tree/master/platform>), o qual já possui um tutorial com instruções que permite replicar totalmente a instalação do ambiente onde estamos desenvolvendo a plataforma. Nela estão todos os passos para poder instalar e carregar a plataforma, incluindo as ferramentas que estamos utilizando (*Django* e *PostgreSQL*). Além disso, atualmente existe no ar uma versão beta que está disponível <http://brascch.picap.cmp.ifsp.edu.br>.

CONCLUSÕES

Este trabalho é parte essencial do projeto supracitado e possui um grande potencial de inovação considerando que atualmente não existe no Brasil uma solução consolidada para o problema em questão. O trabalho está em desenvolvimento, a infraestrutura lógica do servidor web e para armazenamento já estão implementada, assim como o mecanismo para carregamento de arquivos, e formulários para a execução dos modelos de predição está estruturado em fase de testes; por fim atualmente já temos uma visualização carregada na plataforma. Ao final, esperamos que a plataforma possa auxiliar a extrair dos dados, informações relevantes para prevenção da morte neonatal.

AGRADECIMENTOS

Agradecemos o apoio do grupo de pesquisa PICAp-IFSP Campinas; ao apoio financeiro da *Fundação Bill & Melinda Gates* (OPP1201970) e do Conselho Nacional de Desenvolvimento Científico e Tecnológico - CNPq (443774/2018-8), implementado via projeto institucional IFSP (SUAP: 23305.012088.2018-11); ao Programa Institucional de Bolsas de Iniciação Científica e Tecnológica do Instituto Federal de São Paulo - PIBIFSP Campinas (SUAP 23305.012164.2018-98); e à NVIDIA pela doação de uma GPU XP Titan, a ser utilizada nesta pesquisa.

REFERÊNCIAS

- BOOTSTRAP. *Bootstrap*. Disponível em: <<https://getbootstrap.com>>. Acesso: 20 Agosto 2019.
- DJANGO, *Django*. Disponível em: <<https://www.djangoproject.com>>. Acesso: 20 Agosto 2019.
- D3JS. Data-Driven Documents. Disponível em: <<https://d3js.org/>>. Acesso em: 20 Agosto 2019.
- FRIAS, P. G. d. et al. *Utilização das informações vitais para a estimação de indicadores de mortalidade no Brasil: da busca ativa de eventos ao desenvolvimento de métodos*. Cadernos de Saúde Publica, Scielo, v. 33, 00 2017. ISSN 0102- 311X.
- LANSKY, S. et al. *Pesquisa Nascido no Brasil: perfil da mortalidade neonatal e avaliação da assistência à gestante*. Cadernos de Saúde Pública, Scielo, v. 30, p. S192 – S207, 00 2014.
- PYTHON. Python is a programming language that lets you work quickly and integrate systems more effectively. Disponível em: <<https://www.python.org/>>. Acesso: 20 Agosto 2019.
- POSTGRESQL. PostgreSQL: The World's Most Advanced Open Source Relational Database. Disponível em: <<https://www.postgresql.org/>>. Acesso: 20 Agosto 2019.
- REACTJS, React Uma biblioteca JavaScript para criar interfaces de usuário. Disponível em: <<https://reactjs.org/>>. Acesso: 20 Agosto 2019.