



UNIVERSIDADE FEDERAL DE PERNAMBUCO  
GRADUAÇÃO EM CIÊNCIA DA COMPUTAÇÃO  
CENTRO DE INFORMÁTICA  
2016.2

Womens Health Risk Assessment: construindo um modelo de  
aprendizagem de máquina focado nas *features*

**PROPOSTA DE TRABALHO DE GRADUAÇÃO**

Aluna: Déborah Mesquita (dhsm@cin.ufpe.br)

Orientadora: Bernadette Farias Lóscio (bfl@cin.ufpe.br)

Recife, 6 de Setembro de 2016

## 1. Contextualização

Para a maioria da população que habita os países em desenvolvimento a qualidade de vida é bastante precária, em especial para as mulheres. Segundo relatório da Organização Mundial de Saúde (OMS) de 2011, cerca de 820.000 homens e mulheres com idades entre 15 e 24 anos foram recém infectados pelos vírus do HIV nos países em desenvolvimento. Dentre estes infectados, mais de 60% eram mulheres.

Além do HIV e de outras doenças sexualmente transmissíveis, riscos como complicações durante a gravidez e violência física e sexual também assolam grande parte das mulheres. A ênfase na prevenção e no provimento de informações sobre DST e outras infecções relacionadas ao sistema reprodutor feminino foi listada como uma das prioridades para pesquisadores, responsáveis políticos e profissionais de saúde.

A fundação Bill e Melinda Gates disponibilizou um conjunto de dados coletados de pesquisas feitas em 2015 com o objetivo de explorar os desejos, necessidades e comportamentos de meninas e mulheres com relação a sua saúde sexual e reprodutiva em 9 regiões geográficas.

## 2. Objetivos

O objetivo do Trabalho de Graduação é desenvolver um modelo de aprendizagem de máquina capaz de identificar a que segmento de risco pertence cada uma das pacientes que buscaram atendimento em postos de saúde localizados nos países em desenvolvimento. O conjunto de dados consiste em variáveis relacionadas ao perfil socioeconômicos e ao comportamento sexual de mulheres entre 15 e 30 anos.

A pesquisa para o Trabalho será focada em identificar o algoritmo de *multiclass classification* que apresente melhor desempenho para o problema proposto e se dará ênfase ao processo de *feature engineering*, focando extrair informações relevantes a partir do contexto do problema.

O Trabalho é significativo pois foca no processo de *feature engineering*, que consistem em construir *features* a partir dos dados não processados, de forma que facilite o processo de aprendizagem. "Se você possui muitas

*features* independentes e que se correlacionem bem com a classe, o processo de aprendizagem é fácil. Por outro lado, se a classe é uma função complexa das *features* você pode não ser capaz de criar um bom modelo [...] [feature engineering] é a parte onde tipicamente se emprega o maior esforço nos projetos de aprendizagem de máquina. É também usualmente uma das partes mais interessantes, onde a intuição, criatividade e 'magia negra' são tão importantes quanto a parte técnica" (DOMINGOS, 2012).

A pesquisa será conduzida de forma interativa, com as atividades de entendimento e estudo do contexto, análise descritiva dos dados, entendimento do problema e testes de modelos e hipóteses sendo realizadas ao mesmo tempo. Este estilo de pesquisa interativo é importante para o processo de *feature engineering*, onde os *insights* surgem a partir de várias mudanças na natureza das tarefas realizadas.

O trabalho é relevante para o Cin e para a UFPE por abordar um tema relacionado à Ciência dos Dados e assim contribuir para o desenvolvimento desta área de pesquisa no Centro.

### 3. Cronograma

Atividade	Agosto	Setembro	Outubro	Novembro	Dezembro
Análise do contexto					
Análise descritiva dos dados					
Elaboração do modelo de aprendizagem de máquina					
Elaboração do relatório final					
Preparação para a defesa					

### 4. Possíveis Avaliadores

Os possíveis avaliadores deste trabalho de graduação serão:

- Ana Carolina Salgado
- George Darmiton da Cunha Cavalcanti
- Bernadette Farias Lóscio

## 5. Referências

[1] DOMINGOS, Pedro. A Few Useful Things to Know about Machine Learning. Communications of the ACM, Vol. 55 No. 10, Pages 78-87

[2] Womens Health Risk Assessment Competition. Disponível em: <https://gallery.cortanaintelligence.com/Competition/Women-s-Health-Risk-Assessment-1>. Acesso em: 03 de Agosto de 2016.

[3] DAVENPORT, Thomas H; PATIL, D.J . Data Scientist: The Sexiest Job of the 21st Century. Disponível em: <https://hbr.org/2012/10/data-scientist-the-sexiest-job-of-the-21st-century/>. Acesso em: 23 de Agosto de 2016.

[4] LORANG, Noah. Practical skills that practical data scientists need. Disponível em: <https://m.signalvnoise.com/practical-skills-that-practical-data-scientists-needda71e6b93f95#.k4sp3enu1>. Acesso em: 23 de Agosto de 2016.

## 6. Assinaturas

---

Déborah Mesquita

**Orientanda**

---

Bernadette Farias Lóscio

**Orientadora**