



Universidade Federal
do Rio de Janeiro

Escola Politécnica

FRAMEWORK PARA SERVIÇOS DE APRENDIZADO DE MÁQUINA

PEDRO HOLLANDA BOUEKE

Projeto de Graduação apresentado ao Curso de Engenharia de Computação E Informação da Escola Politécnica, Universidade Federal do Rio de Janeiro, como parte dos requisitos necessários à obtenção do título de Engenheiro.

Orientador: Flávio Luis de Mello

Rio de Janeiro
Outubro de 2018

FRAMEWORK PARA SERVIÇOS DE APRENDIZADO DE MÁQUINA

PEDRO HOLLANDA BOUEKE

PROJETO DE GRADUAÇÃO SUBMETIDO AO CORPO DOCENTE DO CURSO
DE ENGENHARIA DE COMPUTAÇÃO E INFORMAÇÃO DA ESCOLA PO-
LITÉCNICA DA UNIVERSIDADE FEDERAL DO RIO DE JANEIRO COMO
PARTE DOS REQUISITOS NECESSÁRIOS PARA A OBTENÇÃO DO GRAU
DE ENGENHEIRO ELETRÔNICO E DE COMPUTAÇÃO

Autor:

PEDRO HOLLANDA BOUEKE

Orientador:

Flávio Luis de Mello, Ph. D.

Examinador:

XXXXXXXXXXXXXXXXXXXXXXXXX

Examinador:

XXXXXXXXXXXXXXXXXXXXXXXXX

Rio de Janeiro
Outubro de 2018

Declaração de Autoria e de Direitos

Eu, *Pedro Hollanda Boueke* CPF 108.243.966.50, autor da monografia *Framework para Serviços de Aprendizado de Máquina*, subscrevo para os devidos fins, as seguintes informações:

1. O autor declara que o trabalho apresentado na disciplina de Projeto de Graduação da Escola Politécnica da UFRJ é de sua autoria, sendo original em forma e conteúdo.
2. Excetua-se do item 1. eventuais transcrições de texto, figuras, tabelas, conceitos e idéias, que identifiquem claramente a fonte original, explicitando as autorizações obtidas dos respectivos proprietários, quando necessárias.
3. O autor permite que a UFRJ, por um prazo indeterminado, efetue em qualquer mídia de divulgação, a publicação do trabalho acadêmico em sua totalidade, ou em parte. Essa autorização não envolve ônus de qualquer natureza à UFRJ, ou aos seus representantes.
4. O autor pode, excepcionalmente, encaminhar à Comissão de Projeto de Graduação, a não divulgação do material, por um prazo máximo de 01 (um) ano, improrrogável, a contar da data de defesa, desde que o pedido seja justificado, e solicitado antecipadamente, por escrito, à Congregação da Escola Politécnica.
5. O autor declara, ainda, ter a capacidade jurídica para a prática do presente ato, assim como ter conhecimento do teor da presente Declaração, estando ciente das sanções e punições legais, no que tange a cópia parcial, ou total, de obra intelectual, o que se configura como violação do direito autoral previsto no Código Penal Brasileiro no art.184 e art.299, bem como na Lei 9.610.
6. O autor é o único responsável pelo conteúdo apresentado nos trabalhos acadêmicos publicados, não cabendo à UFRJ, aos seus representantes, ou ao(s) orientador(es), qualquer responsabilização/ indenização nesse sentido.
7. Por ser verdade, firmo a presente declaração.

Pedro Hollanda Boueke

UNIVERSIDADE FEDERAL DO RIO DE JANEIRO

Escola Politécnica - Departamento de Eletrônica e de Computação

Centro de Tecnologia, bloco H, sala H-217, Cidade Universitária

Rio de Janeiro - RJ CEP 21949-900

Este exemplar é de propriedade da Universidade Federal do Rio de Janeiro, que poderá incluí-lo em base de dados, armazenar em computador, microfilmear ou adotar qualquer forma de arquivamento.

É permitida a menção, reprodução parcial ou integral e a transmissão entre bibliotecas deste trabalho, sem modificação de seu texto, em qualquer meio que esteja ou venha a ser fixado, para pesquisa acadêmica, comentários e citações, desde que sem finalidade comercial e que seja feita a referência bibliográfica completa.

Os conceitos expressos neste trabalho são de responsabilidade do(s) autor(es).

AGRADECIMENTO

Dedico este trabalho ao povo brasileiro que contribuiu de forma significativa à minha formação e estada nesta Universidade. Este projeto é uma pequena forma de retribuir o investimento e confiança em mim depositados.

RESUMO

TODO

Palavras-Chave: trabalho, resumo, interesse, projeto final.

ABSTRACT

TODO

Key-words: word, word, word.

SIGLAS

UFRJ - Universidade Federal do Rio de Janeiro

Conteúdo

1	Introdução	1
1.1	Tema	1
1.2	Delimitação	1
1.3	Justificativa	2
1.4	Objetivos	2
1.5	Metodologia	3
1.6	Descrição	4
2	Informações Adicionais	5
2.1	Citação	5
2.2	Figuras	6
2.3	Tabelas	7
2.4	Numeração de páginas	7
3	Conclusões	9
	Bibliografia	10
A	O que é um apêndice	11
B	Encadernação do Projeto de Graduação	12
C	O que é um anexo	14

Lista de Figuras

2.1	Logotipo do DEL. Fonte: DEL/Poli/UFRJ [1].	6
B.1	Encadernação do projeto de graduação.	13

Lista de Tabelas

2.1	Casos de ataques aos computadores da Intranet. Fonte: DEL/Poli/UFRJ	
	[1].	7

Capítulo 1

Introdução

1.1 Tema

O trabalho apresenta uma solução para uma das questões comumente negligenciadas dentro da área de tecnologias de Aprendizado de Máquina: a implantação e uso de modelos em ambientes não acadêmicos. A proposta executada se reserva à criação de um *framework* para gerência e execução de modelos de Aprendizado de Máquina de forma a automatizar os processos relacionados ao uso de tais modelos, da coleta dos dados a serem analisados, passando por sua execução até a entrega dos resultados.

Nesse trabalho são abordados princípios de Engenharia de Software, algoritmos de Aprendizado de Máquina, *Job Schedulers* e desenvolvimento de aplicações *Web* com a plataforma de desenvolvimento *Django*. Objetiva-se a criação de uma plataforma genérica para a execução periódica de modelos de Aprendizado de Máquina que permita o acompanhamento da execução e resultados por diversos usuários, também facilitando a reconfiguração de agendas e gerência do *workflow* de processamento em ambientes de produção.

1.2 Delimitação

O trabalho apresentado se destina a atender as demandas de agentes de todas as esferas do círculo de usuários de modelos de Aprendizado de Máquina. Enquanto

necessárias execuções recorrentes de modelos, o *framework* desenvolvido apresenta uma solução compatível aos principais contextos em que tais modelos se apresentam. São exceções notáveis a esse conjunto de usuários parte dos desenvolvedores de modelos de Aprendizado de Máquina que, possivelmente, não estejam interessados em praticar execução de seus modelos em ambientes de produção. Também se excluem aqueles que não possuem requisitos técnicos para a execução dos componentes necessários ao sistema, como usuários do sistema operacional Windows, usuários de bancos de dados incompatíveis com a linguagem SQL e usuários impossibilitados de fazer uso da linguagem de programação *Python 3*.

1.3 Justificativa

Enquanto que algoritmos e modelos matemáticos são apresentados frequentemente pelo ambiente acadêmico como soluções aproximadas de problemas reais e complexos condizentes ao que se diz respeito da área de Aprendizado de Máquina, a sua implantação em ambientes não acadêmicos ou de produção poucas vezes é abordada. Dado esse cenário, nota-se que existe espaço e demanda para o estudo e desenvolvimento de novas metodologias e plataformas que ambicionem soluções para as dificuldades envolvidas em trazer um modelo de Aprendizado de Máquina para um ambiente não acadêmico, dentre as quais destacam-se: gerência de execuções e processamento, gerência de arquivos e modelos, visualização e acompanhamento de resultados.

1.4 Objetivos

Objetiva-se o desenvolvimento de um *framework* que permita o agendamento e acompanhamento de execuções de modelos de Aprendizado de Máquina. Uma execução caracteriza-se pela instanciação de um modelo, coleta dos dados a serem processados, execução do modelo sobre os dados coletados e persistência tanto dos resultados quanto das informações relativas à execução. Por acompanhamento, refere-se à possibilidade de usuários do *framework* visualizarem os resultados das execuções, bem como de configurarem todos os aspectos da execução, como a agenda, o modelo e as bases de persistência.

O acima descrito será alcançado com a implementação bem sucedida de um *framework* que permita que todos os detalhes descritos sejam executados, fazendo uso de um sistema de agendamento de tarefas para agendamento de execuções do modelo; uma plataforma de desenvolvimento de aplicações *Web* para permitir o acompanhamento e gerência das execuções; bases de dados que contenham as informações dos usuários do sistema, registros detalhando as etapas das execuções do modelo e os dados a serem processados. Dessa forma, serão objetivos parciais para se alcançar o fim desejado:

1. Desenvolvimento de uma aplicação *Web* caracterizada por:
 - (a) Permitir agendamento de execuções de programas para processamento de modelos de Aprendizado de Máquina.
 - (b) Permitir a visualização de resultados e eventos gerados pelos processos agendados.
 - (c) Se comunicar com bases de dados externas para coleta e persistência de dados.
 - (d) Ser reconfigurável a fim de atender ambientes diversos.
2. Desenvolvimento de uma base de dados relacionais para armazenamento de eventos do *framework*.
3. Desenvolvimento de um programa agendável para execução de modelos de Aprendizado de Máquina.

1.5 Metodologia

Atendendo às expectativas de um projeto de *Software*, iniciou-se o projeto a partir de diagramas e modelos de Engenharia de Software que definem o funcionamento do sistema projetado. Foram elaborados: um diagrama de casos de uso, um diagrama de sequência, cinco diagramas de atividade e um diagrama de relacionamento de entidades.

Com o planejamento teórico concluído, foram definidas as ferramentas de desenvolvimento do sistema. Para controle de versão foi utilizada a ferramenta *Git*, com a manutenção de um repositório para todo o código produzido. Como linguagem de desenvolvimento, foi escolhida a linguagem *Python 3*, que se destaca por ser uma das principais linguagens de programação dentro da comunidade de praticantes de tecnologias de Aprendizado de Máquina [2]. Quanto ao desenvolvimento da aplicação *Web*, foi selecionada a plataforma *Django*, que é caracterizada por depender da mesma linguagem de programação que a escolhida para o *framework* e possuir um sistema de autenticação de usuários embutido. Por fim, quanto aos bancos de dados, foram selecionados: o *SQLite3* para controle de acesso dos usuários, em uma cópia local à aplicação *Web*, o *MySQL* para armazenamento dos eventos produzidos pelo *framework* e, para o banco de coleta e persistência dos dados necessários aos modelos de Aprendizado de Máquina, um banco qualquer que possua compatibilidade com o conector genérico *ODBC*, permitindo consultados *SQL* genéricas.

¡COMO SE DEU O DESENVOLVIMENTO!

1.6 Descrição

Capítulo 2

Informações Adicionais

2.1 Citação

Em um trabalho científico devemos ter sempre a preocupação de fazer referências precisas às idéias, frases ou conclusões de outros autores, isto é, citar a fonte (livro, revista e todo tipo de material produzido gráfica ou eletronicamente) de onde são extraídos esses dados. As citações fundamentam e melhoram a qualidade científica do trabalho, portanto, elas têm a função de oferecer ao leitor condições de comprovar a fonte das quais foram extraídas as idéias, frases ou conclusões, possibilitando-lhe ainda aprofundar o tema/assunto em discussão. Têm ainda como função, acrescentar indicações bibliográficas de reforço ao texto. Veja alguns exemplos:

É neste cenário que "[...] o desafiador cenário globalizado cumpre um papel essencial na formulação das formas de ação." [3].

Segundo Flávio Mello [4], "[...] é claro que o início da atividade geral de formação de atitudes prepara-nos para enfrentar situações atípicas decorrentes do sistema de formação de quadros que corresponde às necessidades [...]".

De acordo com Flávio Mello [5], a certificação de metodologias que nos auxiliam a lidar com o acompanhamento das preferências de consumo causa impacto indireto na reavaliação do orçamento setorial.



Figura 2.1: Logotipo do DEL. Fonte: DEL/Poli/UFRJ [1].

Por outro lado, a mobilidade dos capitais internacionais garante a contribuição de um grupo importante na determinação dos índices pretendidos. Percebemos, cada vez mais, que o novo modelo estrutural aqui preconizado deve passar por modificações independentemente das condições inegavelmente apropriadas [6].

Além disto, a expressão latina **apud** que significa: citado por, conforme, segundo é utilizada quando se faz referência a uma fonte secundária. Suponha que você teve acesso ao conteúdo do texto de Fulado através do trabalho Beltrano: Não obstante, a competitividade nas transações comerciais prepara-nos para enfrentar situações atípicas decorrentes do fluxo de informações. A determinação clara de objetivos garante a contribuição de um grupo importante na determinação dos relacionamentos verticais entre as hierarquias. O incentivo ao avanço tecnológico, assim como a mobilidade dos capitais internacionais desafia a capacidade de equalização de alternativas às soluções ortodoxas (Fulano [7] apud Beltrano [8]).

2.2 Figuras

Figuras (organogramas, fluxogramas, esquemas, desenhos, fotografias, gráficos, mapas, plantas e outros) constituem unidade autônoma e explicam, ou complementam visualmente o texto, portanto, devem ser inseridas o mais próximo possível do texto a que se referem. Sua identificação deverá aparecer na parte inferior precedida da palavra designativa (figura), seguida de seu número de ordem de ocorrência, do respectivo título e/ou legenda e da fonte, se necessário, tal como na Figura 2.1.

2.3 Tabelas

As tabelas são elementos demonstrativos de síntese que apresentam informações tratadas estatisticamente constituindo uma unidade autônoma. Em sua apresentação deve ser observado: (1) o título deverá ser colocado na parte inferior, precedido da palavra Tabela e de seu número de ordem; (2) as fontes e eventuais notas aparecem em seu rodapé, após o fechamento, utilizando-se o tamanho 10; (3) Devem ser inseridas o mais próximo possível do trecho a que se referem, tal como a Tabela 2.3.

Tabela 2.1: Casos de ataques aos computadores da Intranet. Fonte: DEL/Poli/UFRJ [1].

Número IP	Ataques	Ataques bem sucedidos
192.168.0.120	54	1
192.168.0.123	36	2
192.168.0.129	25	4
192.168.0.130	16	0
192.168.0.141	29	3
Total	160	10

2.4 Numeração de páginas

O aluno deve observar atentamente a numeração de páginas de seu projeto. A primeira parte deste modelo de projeto final, composta pela dedicatória, agradecimento, resumo, abstract, siglas, sumário, lista de figuras e lista de tabelas, é numerada sequencialmente utilizando algarismos romanos minúsculos. As demais folhas, descritas na segunda parte deste modelo, são numeradas sequencialmente utilizando algarismos arábicos.

Contudo, exclusivamente para a segunda parte do modelo de projeto, é permitida uma numeração alternativa na qual o aluno poderá numerar as páginas por capítulo. Por exemplo, a primeira página deste Capítulo 2 - Informações Adicionais, poderia ser escrita como 2.1. Além disto, a página seguinte seria 2.2 e a presente página poderia ser escrita como 2.3. A página do Apêndice A - O que é um apêndice,

poderia ser escrita como A.1, enquanto que a primeira página do apêndice B seria B.1. Neste caso alternativo específico, a Bibliografia na deverá conter numeração.

Capítulo 3

Conclusões

Tratam-se das considerações finais do trabalho, mostrando que os objetivos foram cumpridos e enfatizando as descobertas feitas durante o projeto. Em geral reserva-se um ou dois parágrafos para sugerir trabalhos futuros.

Observe que neste modelo a conclusão é numerada pelo numeral 3, mas o projeto não tem a obrigatoriedade de possuir apenas 3 capítulos. Alias, espera-se que tenha mais que isso.

Bibliografia

- [1] MEYER, D. E., KIERAS, D. E., *Título da nota tecnica*, Report TR-97/ONR-EPIC-08, Department of Psychology, Electrical Engineering & computer Science Department, University of Michigan, 1997.
- [2] THE STACK OVERFLOW NETWORK, “Stack Overflow Annual Developer Survey”, <https://insights.stackoverflow.com/survey>, 2018, [Data File].
- [3] LEWIS, H. R., PAPADIMITRIOU, C. H., *Título do livro*. Porto Alegre, Bookman, 2004.
- [4] NETTLE, P., “Título da URL”, <http://www.brainvoyager.com/>, 2006, (Acesso em 05 Janeiro 2006).
- [5] HEEGER, D., RESS, D., “Título do artigo”, *Nature Reviews/Neuroscience*, v. 3, pp. 142–151, 2002.
- [6] CANGUILHEM, G., “Título do artigo em proceeding”. In: *Georges Canguilhem - Philosophe, historien des sciences: Actes du Colloque*, pp. 11–33, Paris, 1993.
- [7] FULANO, J., *Título da dissertação de mestrado*. M.Sc. dissertation, Universidade Federal do Rio de Janeiro, Julho 2003.
- [8] BELTRANO, M., *Título do livro onde se consultou um capítulo*, chapter Membrane potential and action potential, New York, Academic Press, pp. 129–154, 1999.

Apêndice A

O que é um apêndice

Elemento que consiste em um texto ou documento elaborado pelo autor, com o intuito de complementar sua argumentação, sem prejuízo do trabalho. São identificados por letras maiúsculas consecutivas e pelos respectivos títulos.

Apêndice B

Encadernação do Projeto de Graduação

Número	<p>UNIVERSIDADE FEDERAL DO RIO DE JANEIRO</p> <p>Escola Politécnica</p> <p>Departamento de Eletrônica e de Computação</p>
Nome do Aluno	
Título do Projeto*	<p>Título do Projeto</p> <p>Nome do Aluno</p>
Ano	<p>Projeto de Graduação</p> <p>Mês / Ano</p>

*** Título resumido caso necessário**
Capa na cor preta, inscrições em dourado

Figura B.1: Encadernação do projeto de graduação.

Apêndice C

O que é um anexo

Documentação não elaborada pelo autor, ou elaborada pelo autor mas constituindo parte de outro projeto.