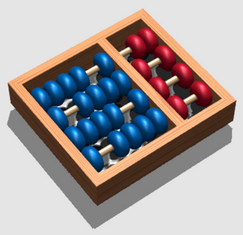

Workflow autonômico para aplicações de *Machine Learning* utilizando métricas de *Fairness*

Autor: Thales Eduardo Nazatto (tenazatto@gmail.com)
Orientador(es): Cecília Mary Fischer Rubira (cmrubira@ic.unicamp.br), Leonardo Montecchi (leonardo.montecchi@ntnu.no)

Palavras-chave: *Workflow*, *Machine Learning*, Inteligência Artificial, Computação Autônoma, Métricas de *Fairness*

1 Introdução

O uso de Inteligência Artificial (IA) envolvendo grandes volumes de dados vem crescendo conforme nossa sociedade migra processos manuais de trabalho para soluções digitais e necessita de tomadas de decisão mais rápidas e assertivas, mas, devido a barreiras éticas e legais, métricas usadas inicialmente para definir a eficácia de um algoritmo se mostraram limitadas para medir vieses que refletem a sociedade de maneira que não era esperada pelos desenvolvedores da solução. Para resolver tal problema, novos algoritmos foram desenvolvidos e um novo conjunto de métricas, denominado como métricas de Fairness, é utilizado para determinar um equilíbrio entre grupos que sofrem discriminações. Entretanto, novos problemas surgem com a introdução deste novo conjunto de algoritmos e métricas, como a piora nas métricas tradicionais de avaliação e aumento de combinações de algoritmos utilizados no processo, aumentando a complexidade da análise do Cientista de Dados para obter modelos de forma otimizada.

2 Objetivo

O objetivo deste trabalho é desenvolver uma estrutura de *Workflow* para aplicações de *Machine Learning* que seja completamente autônoma, por três fatores principais:

- Facilitar a criação de modelos justos e confiáveis com a automatização da escolha dos algoritmos, cuja complexidade aumenta com a escolha dos algoritmos a serem utilizados e suas execuções nas etapas corretas do processo, onde eles foram escolhidos para atuar.
- Estabelecer um balanceamento entre métricas para avaliar bons modelos com métricas para avaliar modelos justos.
- Considerar proveniência de dados como requisito no design de uma solução de IA, e como uma alternativa a *Explainable AI* através da utilização de metadados.

3 Metodologia

Foi desenvolvido um sistema, que pode ser dividido em 4 etapas principais:

- **Engenharia de dados:** Etapa criada com o objetivo de simular processos de transformação e limpeza de dados.
- ***Workflow* de IA:** Etapa para execução de um *Workflow* que simula o desenvolvimento de uma aplicação automatizada de IA, desde uma categorização dos dados mais específica do que na etapa anterior, passando pelo algoritmo utilizado e finalizando obtendo métricas para determinar qualidade do resultado final.
- **Autonomia do *Workflow*:** Etapa que automatiza todas as etapas do *Workflow* através de um componente, com o objetivo de evitar com que perca-se tempo em execuções manuais que podem demorar dependendo do algoritimo e do conjunto de dados utilizado.
- **Interface Humano-Computador:** Etapa criada com o objetivo de simular a etapa anterior, porém de modo a proporcionar uma experiência de usuário mais simples e intuitiva, onde sua integração com as outras etapas é mostrada na Figura 1. É dividida em duas partes:
  - **Frontend:** Parte visual, exibida em um navegador.
  - **Backend:** Parte onde o Frontend se comunica para obter os dados para auxiliar a montagem do visual e executa o componente utilizado na etapa de autonomia, de forma que corresponda a configurações utilizadas por ela.

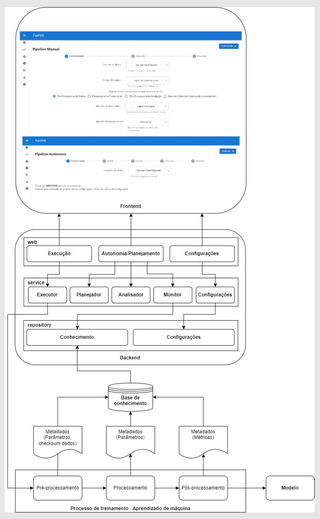

Figure 1: Comunicação entre interface, autonomia e *Workflow* de IA

Para o desenvolvimento do *workflow*, será utilizada a arquitetura *Pipe-and-Filter*. Para a autonomia deste, será criado um componente utilizando a arquitetura MAPE-K [IBM(2005)] para analisar uma base de conhecimento e prover o melhor pipeline seguindo regras pré-determinadas. Para a interface, ela foi criada nos moldes de uma aplicação web. O código deste desenvolvimento foi disponibilizado no GitHub (`https://github.com/tenazatto/MsC`) para avaliação e testes em estudos posteriores.

4 Resultados

Foram realizados estudos de caso para verificar a viabilidade da arquitetura MAPE-K na autonomia e a capacidade de evolução para o uso de novos algoritmos e conjuntos de dados. Foi atribuída uma pontuação de 0 a 1000 no cálculo realizado pela análise.

	Workflow			Pontuação		
Atributo protegido	Pré-processamento	Treinamento	Pós-processamento	Performance	Fairness	**Geral**
Idade	Nenhum	Regressão Logística	Equalized Odds	968	860	**914**
Nacionalidade	Nenhum	Random Forest	Calibrated Equalized Odds	902	922	**912**
Nacionalidade	Nenhum	Gradient Boosting	Calibrated Equalized Odds	870	925	**898**
Idade	Nenhum	Gradient Boosting	Equalized Odds	927	862	**894**
Idade	Reweighing	Gradient Boosting	Nenhum	804	931	**868**

É possível ver que há uma diversidade entre os parâmetros do workflow e suas pontuações entre métricas de performance e métricas de *Fairness*, a arquitetura MAPE-K se mostra viável para efetuar um balanceamento. Com diferentes definições de pesos para as métricas, é possível calibrar qual a melhor escolha dependendo do contexto do problema.
Como exemplo para evolução, foram necessárias modificações para adicionar um novo conjunto de dados como opção no *Workflow*. Estas foram contadas de acordo com seus *commits* realizados no repositório e exibidos na tabela abaixo:

Parte do Sistema	Linhas alteradas	Total de linhas	Arquivos alterados	Total de arquivos	% linhas alteradas	% arquivos alterados
Engenharia de Dados	122	277	2	3	**44,04%**	**66,67%**
Workflow de IA	76	1982	5	38	**3,84%**	**13,16%**
Autonomia do *Workflow*	0	457	0	10	**0,00%**	**0,00%**
Interface Humano-Computador (*Frontend*)	13	2905	2	14	**0,45%**	**14,29%**
Interface Humano-Computador (*Backend*)	4	432	1	7	**0,93%**	**14,29%**
TOTAL	**215**	**6053**	**10**	**72**	**3,55%**	**13,89%**

Destas modificações, percebe-se que a grande maioria delas está relacionada a parte de engenharia de dados. Mesmo no *Workflow* de IA, a maior parte dessas linhas não está relacionada a sua estrutura e sim a uma etapa de adaptação dos dados ao atributo protegido, necessário para algoritmos dedicados a reduzir viés nos dados. A etapa de autonomia e a Interface Humano-Computador exigiram poucas ou nenhuma alteração e em poucos arquivos, comprovando que a modularização das arquiteturas escolhidas facilita a manutenção do sistema e que a autonomia não é afetada por evoluções no *Workflow*.

5 Conclusões

Esta pesquisa mostra que a escolha da arquitetura *Pipe-and-Filter* se mostra favorável para o desenvolvimento de um *workflow* para aplicações envolvendo IA, permitindo que ele seja modular e que sejam feitas evoluções sem exigir grandes esforços. O uso da arquitetura MAPE-K também se mostrou favorável, permitindo diversos resultados para diferentes contextos de problema e uma simplificação da análise realizada pelo Cientista de Dados, podendo resultar em economia de tempo. Ela também possibilitou um balanço entre performance e justiça através da adição de pesos para cada métrica na parte de análise. Embora os pesos não sejam parte da arquitetura, a divisão presente na arquitetura permite que o desenvolvimento seja pensado de maneira mais clara.
Embora a proveniência de dados não teve o mesmo efeito da aplicação de um método feito para *Explainable AI*, a obtenção de metadados do *workflow* se mostrou essencial para alimentar o componente baseado na arquitetura MAPE-K e possibilitou a análise e tomadas de decisão baseadas em dados. É possível realizar análises mais detalhadas conforme novas execuções forem realizadas, consequentemente podendo resultar em melhores escolhas de algoritmos para contextos distintos, e até considerar novas aplicações de IA caso o volume de dados seja consideravelmente grande. Posteriormente, é possível também rever quais dados podem ser adicionados para que o objetivo de *Explainable AI* possa ser alcançado de maneira satisfatória.

Referências

[IBM(2005)] An architectural blueprint for autonomic computing. Technical report, IBM, 2005. URL `https://www-03.ibm.com/autonomic/pdfs/AC%20Blueprint%20White%20Paper%20V7.pdf`.