

Trabalho prático de Mineração de Dados e Aprendizado de Máquina

Auditorias em Planos de Saúde

Objeto: Autorização de procedimentos de uma operadora de Planos de Saúde.

Conceitos iniciais

Operadora: Empresa habilitada a comercializar planos de saúde, com o objetivo de intermediar o acesso a serviços assistenciais.

Prestador: Empresa ou pessoa que a operadora contrata para prestar assistência em saúde para os seus clientes.

Beneficiário: Pessoa física que contrata a operadora (cliente final).

Contexto

O principal objetivo da operadora é controlar os gastos gerados pelos beneficiários, garantindo que eles realizem apenas os procedimentos contratualizados e que sejam pertinentes para a manutenção da saúde. Para isso, a operadora exige que todos os atos assistenciais sejam previamente autorizados.

Em resumo, o prestador questiona a operadora se o beneficiário está apto a realizar o procedimento. O atendimento só deve continuar se a resposta for positiva. Todas as solicitações são analisadas pelo sistema informatizado da operadora. Após a análise automática, o sistema define se já consegue dar a resposta final (sim ou não) ou depende da análise de um auditor (funcionário da operadora).

Os auditores analisam todos os artefatos da requisição (justificativa, anexos, histórico do beneficiário etc.) e respondem se o procedimento está autorizado ou não.

Problema

Os auditores são profissionais especializados (médicos e enfermeiros) que geram custos elevados para a operadora. Quanto maior o número de requisições encaminhadas para auditoria, maior será o gasto administrativo.

Os sistemas atuais se limitam a analisar as regras parametrizadas (se/então), não conseguem analisar o comportamento dos auditores para prever desfechos prováveis.

Proposta

Gerar modelos capazes de combinar as variáveis de uma requisição e, baseada no comportamento histórico dos auditores, consiga automatizar a análise de novas solicitações.

Realização do trabalho

- Este trabalho será realizado como uma competição no Kaggle. Os conjuntos de treinamento e teste estão disponíveis em <https://www.kaggle.com/competitions/auditoria-bia-ufg>
- Participação de grupos de no máximo 4 alunos, já definidos em aula.
- As atividades deverão ser realizadas **presencialmente** na sala e no horário da **aula de Mineração de Dados, e também extraclasse** em encontros definidos pelos próprios componentes dos grupos.
- Cada grupo deverá ter um notebook no Google Colab compartilhado entre os membros do grupo e o professor.
- Todas as atividades do processo (desde a análise estatística dos dados até a geração e avaliação dos modelos) devem estar documentadas no notebook.

Avaliação

O trabalho será acompanhado pelo professor durante as aulas e a avaliação se dará por:

- observação dos grupos trabalhando nas atividades
- questionamentos
- uso correto e apropriado das técnicas, além de estarem bem documentadas no notebook
- pesquisa e aplicação de técnicas não vistas em sala de aula
- presença de todos os componentes do grupo.

O Kaggle irá avaliar automaticamente a performance dos modelos submetidos. A métrica de avaliação será a **F1-score**. Os detalhes sobre a submissão estão na página da competição. Além disso, devem ser observados os seguintes pontos:

- haverá limiares para a performance mínima, boa e muito boa.
- o modelo final submetido na plataforma deverá atingir pelo menos a performance mínima.

Cronograma

Data	Entrega mínima
Aula da semana 28/03	Descrição e análise estatísticas dos dados (sumários, visualizações...). Discussão sobre as tarefas de limpeza do dados (dados irrelevantes, faltantes, duplicados, inconsistentes, outliers...).
Aula da semana 04/04	Limpeza de dados feita. Discussão sobre as tarefas de transformação, seleção e criação de atributos (normalização, padronização, discretização, conversões de tipos de dados...).
Aula da semana 11/04	Conjunto de dados processado. Discussão sobre os algoritmos e métodos de mineração de dados a serem aplicados (algoritmos de aprendizado, como será avaliada a expectativa do erro de generalização, verificação de outras métricas de avaliação de modelos).
15/04	Submissão final no Kaggle.

