



De 11 a 25 de novembro de 2023

Ciências Exatas e da Terra
FEMIC JOVEM

Rafael Moreira de Melo

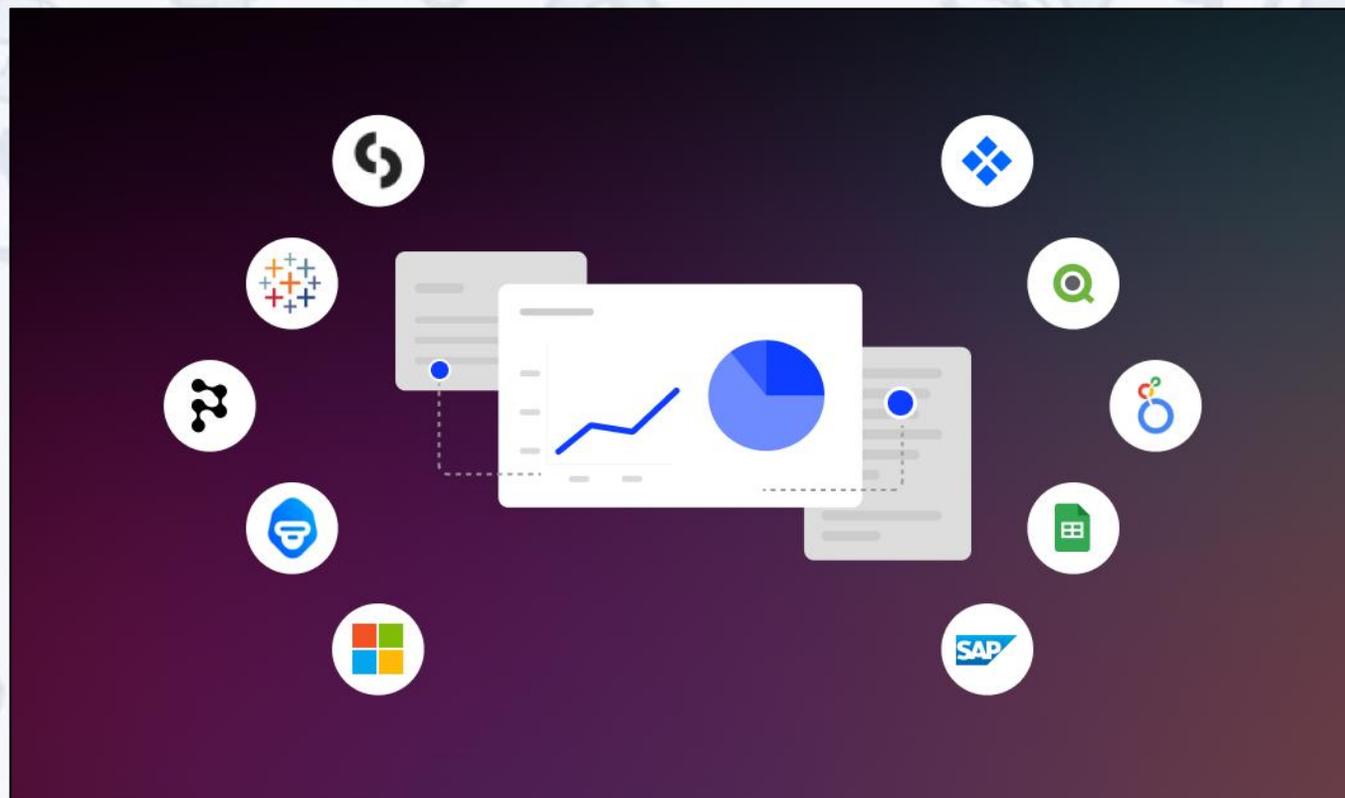
Samuel Machado

Orientador: Jair Rodrigues de Andrade

Escola Estadual Gregoriano Canedo

Monte Carmelo, Minas Gerais e Brasil

APLICANDO O MODELO PREDITIVO DE APRENDIZADO DE MÁQUINA PARA O DIAGNÓSTICO DO DIABETES TIPO 2 COM LINGUAGEM R



Digite aqui um e-mail de contato

Apresentação



- A ciência de dados é o estudo dos dados para extrair insights e desenvolver modelos de dados utilizando o aprendizado de máquina visando prever algo com base em dados, ou seja, um conjunto de informações, como no projeto, utilizamos dados, como insulina, glicose, e outros, sobre pacientes diagnosticados com diabetes para reconhecer padrões e montar modelos utilizados para prever se uma pessoa tem ou não diabetes.
- Atualmente é constantemente observado a mudança no comportamento da população, com o crescimento tecnológico, período pandêmico e outros fatores, em que as pessoas estão se tornando cada vez mais sedentárias e com isso agravando situações problemáticas de saúde. E isso afetou fortemente no risco das pessoas obterem um quadro de diabetes. Dito isso, iremos desenvolver uma pesquisa com o nosso grupo de iniciação científica para realizar um estudo e criação de modelos de aprendizado de máquina utilizando um conjunto de dados coletado pelo Instituto Nacional de Diabetes e Doenças Digestivas e Renais dos Estados Unidos, lembrando que o objetivo do aprendizado de máquina não é substituir os profissionais da saúde, e sim melhorar o atendimento da população quanto ao diagnóstico dessa enfermidade.

Objetivos



- O projeto de ciência de dados consiste em explorar a inteligência artificial, por meio dos algoritmos Regressão Logística e Floresta Aleatória, visando criar dois modelos de aprendizado de máquina para prever o diagnóstico do diabetes tipo 2.
- Promover discussões sobre metodologia de pesquisa na área de ciência da natureza, matemática e suas tecnologias utilizando de banco de dados explorados e publicados.
- Mostrar a importância do campo da ciência de dados, o conhecimento sobre os dados para desenvolver uma multidisciplinaridade nos diversos campos do conhecimento, além do notado crescimento da área nos últimos tempos.

Metodologia



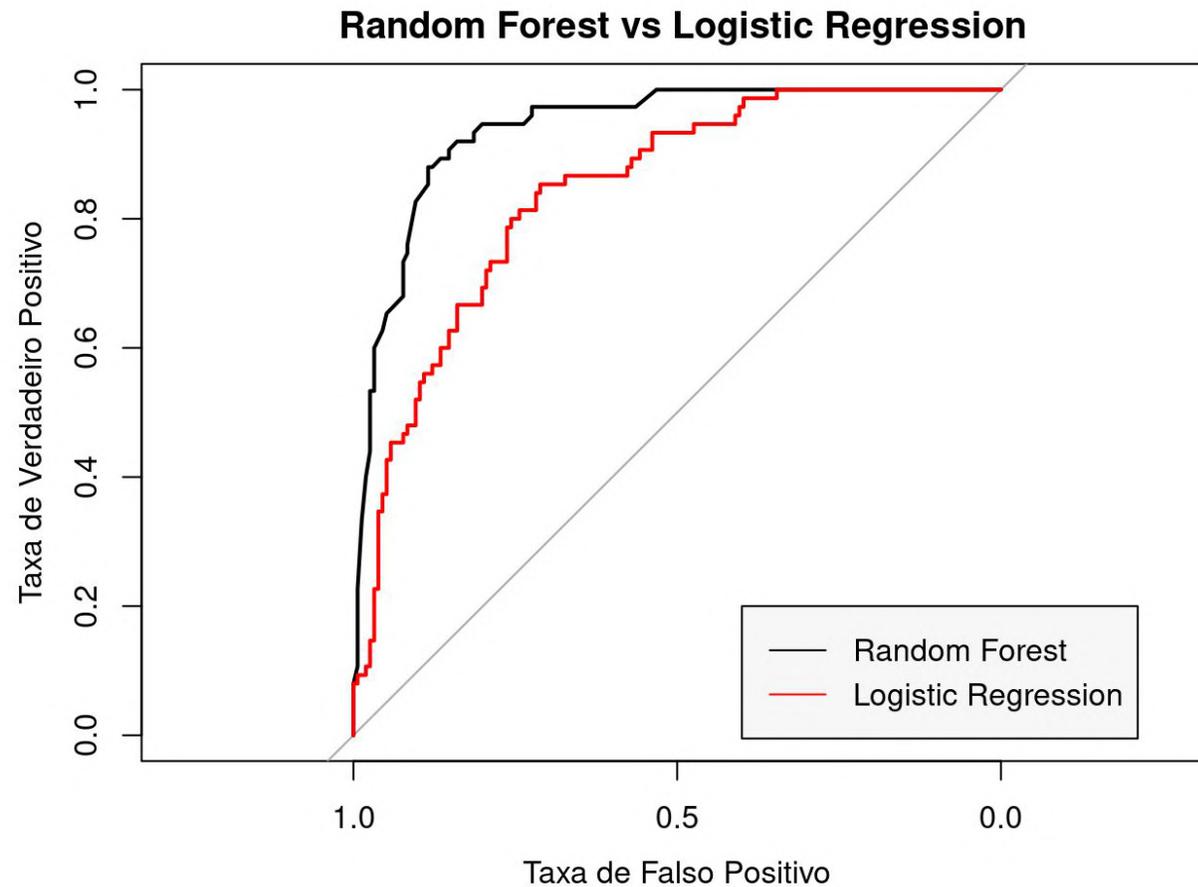
Ciclo de Análise de Dados



Metodologia



- Comparação do modelo preditivo utilizando Regressão Logística e por último e otimizado Floresta Aleatória



Resultados alcançados



- O primeiro modelo criado, utilizando Regressão Logística, apresentou uma precisão de aproximadamente 77%, sendo o mínimo do objetivo 78%.

```
## Confusion Matrix and Statistics
##
##           Reference
## Prediction    0    1
##           0 133  30
##           1  23  45
##
##           Accuracy : 0.7706
##           95% CI : (0.7109, 0.8232)
##           No Information Rate : 0.6753
##           P-Value [Acc > NIR] : 0.000957
##
```

Resultados alcançados



- O segundo modelo, utilizando o algoritmo Floresta Aleatória, conseguiu atingir o objetivo de no mínimo 78%, apresentando uma precisão de 88% com base nos dados de teste.

```
## Confusion Matrix and Statistics
##
##           Reference
## Prediction    0    1
##           0 138    9
##           1  18   66
##
##           Accuracy : 0.8831
##           95% CI : (0.8345, 0.9215)
##           No Information Rate : 0.6753
##           P-Value [Acc > NIR] : 1.82e-13
##
```

Aplicabilidade dos resultados no cotidiano da sociedade



- O algoritmo **Floresta Aleatória** é um método utilizado para criar um modelo a partir de um conjunto de observações, usado na predição de valores tomados por uma variável categórica, a partir de uma série de variáveis explicativas contínuas.
- A Floresta Aleatória é utilizada em algumas áreas, como:
- Saúde: melhorando o atendimento dos pacientes, por exemplo, com um melhor gerenciamento do tempo e das esperas.
- Instituições Financeiras: pode identificar grupos sujeitos a risco para a subscrição de crédito.
- Econometria: permite explicar uma variável discreta, como por exemplo as intenções de voto em atos eleitorais.
- Tecnologia: atualmente vemos os algoritmos tomando conta da tecnologia, por exemplo, quando um aplicativo reconhece, por meio do seu comportamento de uso, suas preferências de músicas, ou filmes, e depois ele irá recomendar algo parecido.

Criatividade e inovação



- O campo dos dados está tendo uma grande alta atualmente, já que por meio dele podemos reconhecer padrões, pontos fora da curva e tirar algumas ideias disso para resolver problemas de negócio.
- E com essa onda da tecnologia e inteligência artificial, também podemos falar do aprendizado de máquina, que utiliza dados, um conjunto de observações, para treinar modelos preditivos capazes de prever algo com base nesses dados. Tivemos no projeto o treinamento de dois desses algoritmos, a Regressão Logística e a Floresta Aleatória.

Considerações finais



- O rastreamento do diabetes tipo 2 por meio de modelos preditivos treinados com o algoritmo Floresta Aleatória, presentes no Aprendizado de Máquina, que é um subcampo da Inteligência Artificial, permite pré-diagnosticar um paciente com diabetes a partir de uma combinação de dados com características comuns em diabéticos. Com este projeto queremos ressaltar a importância de realizar uma análise exploratória correta e daí selecionar o algoritmo correto e assim criar um modelo preditivo correto, além de observar os valores faltantes, os valores discrepantes e também a correlação entre as variáveis do conjunto de dados, já que isso tudo afeta significativamente a precisão do modelo.

Agradecimentos

Gostaríamos de agradecer ao nosso orientador **Jair Rodrigues de Andrade** pela oportunidade de aprender mais sobre **Ciência de Dados** e as aplicações do **Aprendizado de Máquina**. Nossos agradecimentos também a nossa **Escola Estadual Gregoriano Canedo**, que nos incentivou e nos deu o espaço necessário para o desenvolvimento deste trabalho. E também pelo apoio do **CNPq**.



7ª Feira Mineira de Iniciação Científica

De 11 a 25 de novembro de 2023

Realização



Associação Mineira de
Pesquisa e Iniciação Científica



Apoiadores



MINISTÉRIO DA
CIÊNCIA, TECNOLOGIA
E INOVAÇÃO

