

**Pós-graduação em Inteligência Artificial**

**Redes Neurais e Deep Learning**

**Diabetes**

**Prof: Tarcísio Pontes**

**Marleson Filipe / Rodrigo Alves**

**Julho 2019**

1. **Introdução**

O diabetes é uma doença comum no território brasileiro, e atinge uma grande quantidade de pessoas devido às ausências de uma alimentação saudável, prática de exercícios físicos, consumo de álcool e cigarro.

Junto com o avanço da área de Inteligência Artificial, focado na resolução de problemas complexos no processamento de dados, enxergamos nas Redes Neurais a possibilidade de analisar grandes massas de dados, visando estimar a probabilidade do desenvolvimento da doença pelo indivíduo baseando-se em questionários de saúde.

1. **Objetivo**

Identificar hábitos/características que, associadas, podem ser determinantes para o diagnóstico de pessoas que possuem, ou, poderão desenvolver, diabetes.

1. **Descrição do problema**

O diabetes é uma doença crônica onde o pâncreas não produz insulina suficiente ou quando o corpo não consegue utilizá-la de maneira eficaz. A insulina é o hormônio que regula a glicose no sangue e é fundamental para manutenção do bem-estar do organismo, que precisa da energia dela para funcionar. Altas taxas de glicose podem levar a complicações no coração, nas artérias, nos olhos, nos rins e nos nervos. Em casos mais graves, o diabetes pode levar à morte. De acordo com a [Sociedade Brasileira de Diabetes](https://www.diabetes.org.br/), existem atualmente, no Brasil, mais de 13 milhões de pessoas vivendo com a doença, o que representa 6,9% da população nacional. A melhor forma de prevenir é praticando atividades físicas regularmente, mantendo uma alimentação saudável e evitando consumo de álcool, tabaco e outras drogas.

**Os principais tipos de diabetes são:**

**Diabetes Tipo 1:** a causa desse tipo de diabetes ainda é desconhecida e a melhor forma de preveni-la é com práticas de vida saudáveis (alimentação, atividades físicas e evitando álcool, tabaco e outras drogas).

**Diabetes Tipo 2:** ocorre quando o corpo não aproveita adequadamente a insulina produzida. Esse tipo de diabetes está diretamente relacionado ao sobrepeso, sedentarismo e hábitos alimentares inadequados.

**Diabetes Latente Autoimune do Adulto (LADA):** atinge basicamente os adultos e representa um agravamento do diabetes tipo 2.

**Diabetes gestacional:** ocorre temporariamente durante a gravidez. As taxas de açúcar no sangue ficam acima do normal, mas ainda abaixo do valor para ser classificada como diabetes tipo 2. Toda gestante deve fazer o exame de diabetes, regularmente, durante o pré-natal. **Mulheres com a doença têm maior risco de complicações durante a gravidez e o parto.**

1. **Resumo sobre Redes Neurais com foco nas técnicas utilizadas**

As redes neurais compreendem procedimentos computacionais que envolvem o desenvolvimento de estruturas matemáticas com habilidade de aprendizado. Representam o esforço de investigações acadêmicas para implementar computacionalmente, a maneira pela qual o cérebro humano funciona. São programas que implementam detecções sofisticadas de padrões e algoritmos de aprendizado de máquina, para construir modelos, principalmente, de prognóstico de grandes bancos de dados históricos. Está baseada nos conceitos de como um cérebro humano está organizado e como ele aprende. Existem duas estruturas principais: (1) O nó, que corresponde ao neurônio; (2) O link, que corresponde às conexões entre neurônios.

Neste estudo, aplicamos o conhecimento supervisionado(orientado), onde existe um valor para ser prognosticado, uma classe a ser atribuída aos registros ou um determinado relacionamento para ser explorado. Existe apenas uma vaga idéia do que se está procurando. Os passos para aplicação da busca de conhecimento direta são:

• Identificar as fontes dos dados selecionados para mineração;

• Preparar os dados para análise;

• Construir e treinar o modelo computacional;

• Avaliar o modelo computacional.

Existem diversos algoritmos para redes neurais artificiais, mas o utilizado neste projeto é o MLP (MultiLayer Perceptron ou Perceptron Multicamadas). Para que uma rede dessas funcione, é preciso treiná-la. É como ensinar a uma criança o beabá. O treinamento de uma rede MLP insere-se no contexto de aprendizado de máquina supervisionado, em que cada amostra de dados utilizada apresenta um rótulo informando a que classificação ela se encaixa. Por exemplo, uma imagem de um cachorro contém um rótulo informando que aquilo é um cachorro. Assim, a ideia geral é fazer com que a rede aprenda os padrões referentes a cada tipo de coisa (cada *classe*), assim, quando uma amostra desconhecida for fornecida à rede, ela seja capaz de estabelecer a qual classe tal amostra pertence.

1. **Experimentos**

Experimento Rodrigo:

A base utilizada foi a de diabetes, onde o número de atributos foram reduzidos de 113 para 69. Foram utilizados 31023 exemplos para realizar o experimento, onde 70% foi utilizado para treinamento e 30% para validação usando ainda a separação do tipo KFold. Os atributos foram escolhidos levando em consideração hábitos que tivesse relação direta com o surgimento da doença, como por exemplo, o consumo excessivo de alimento com nível elevado de açúcar como refrigerantes, bolos, doces, consumo de bebida alcoólica e cigarros a base de nicotina, e a falta da prática de exercícios físicos e de controle de peso.

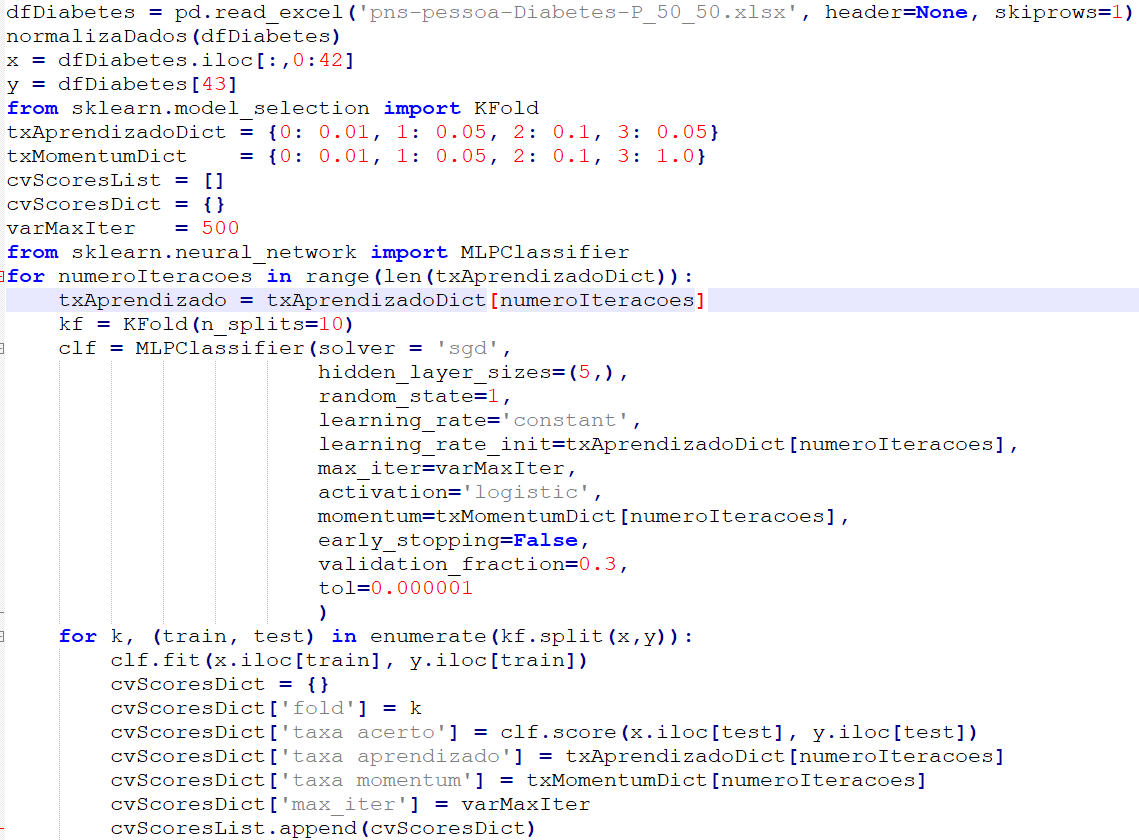
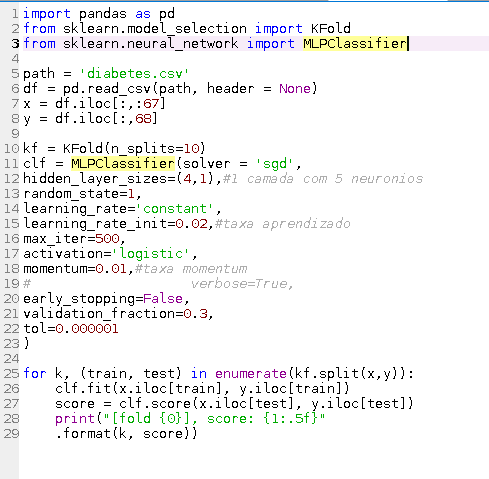
Todos os atributos foram normalizados e convertidos em valores de 0 e 1. Esse processo foi realizado utilizando a ferramenta de datawarehouse, Pentaho.

Experimento Filipe:

A base utilizada foi a de diabetes, onde o número de atributos foram reduzidos de 112 para 43, com base no levantamento de informações relevantes no site <http://www.saude.gov.br/saude-de-a-z/diabetes>, totalizando os atributos em: 1 numérico, 35 categóricos e 3 booleanos. Após a redução e limpeza dos atributos, os dados foram convertidos em valores de 0 a 1 através de um algoritmo em Python.

Os testes iniciais foram realizados na base *pns\_pessoa\_Diabetes\_P10p*, onde contém 10% de dados da base completa, atingindo uma taxa de acerto de 94% através de técnicas como kFold 10, e MLP. Em seguida, foi aplicado os mesmos testes, com seus respectivos pesos, na base completa *pns\_pessoa\_Diabetes\_P*, atingindo uma taxa de 94%. Verificamos então, que poderíamos estar em um cenário de *Overfiting,* então em seguida a base foi organizada a fim de deixar os dados com os labels em proporções equilibradas.

1. **Código**

****

1. **Conclusão**

Redes neurais são um exemplo de que é possível construir máquinas capazes de aprender, de errar, de melhorar e principalmente, que não são presos a programações que podem falhar. O que limita uma máquina dessas é sua própria experiência

1. **Referências**

<https://www.diabetes.org.br/>

<http://www.saude.gov.br/saude-de-a-z/diabetes>

<https://saude.ig.com.br/alimentacao-bemestar/2012-11-08/feijao-e-outras-leguminosas-melhoram-a-pressao-e-o-diabetes-aponta-estudo.html>

<https://www.terra.com.br/vida-e-estilo/saude/doencas-e-tratamentos/carne-vermelha-pode-aumentar-risco-de-diabetes-tipo-2,77583f04c2f27310VgnCLD100000bbcceb0aRCRD.html>

<https://www.minhavida.com.br/alimentacao/materias/17458-pessoas-com-diabetes-precisam-controlar-o-consumo-de-gorduras>