**Universidade Presbiteriana Mackenzie**

**Integrantes:**

**Bruno Zovaro Nascimento - 10424880**

**Douglas Novaes Dias – 14023666**

**Milan Mirco Moraes Mazur – 10363757**

**Paulo Cesar Masson Junior - 10416023**

**PROJETO APLICADO II**

**SUBCONJUNTO DE PREVISÃO DE IDADE**

**São Paulo**

**2024**

**SUMÁRIO – Aplicando Conhecimento – Aula 2**

1 Definição da linguagem de programação usada no projeto

2 Análise exploratória da base de dados escolhida

3 Tratamento da base de dados (Preparação e treinamento)

4 Definição e descrição das bases teóricas dos métodos

5 Definição e descrição de como será calculada a acurácia

6 Base de Dados/ Repositório do Github

7 Cronograma de Atividades

1. **Definição da linguagem de programação usada no projeto**

A linguagem de programação é Python. As bibliotecas a serem usadas incluem pandas, seaborn, fetch\_ucirepo, sklearn.

Texto

Descrição gerada automaticamente

1. **Análise exploratória da base de dados escolhida.**

Head:

Tela de computador

Descrição gerada automaticamente

Describe:

Interface gráfica do usuário, Texto

Descrição gerada automaticamente

1. **Tratamento da base de dados (Preparação e treinamento).**

Preparação 1:

Tela de computador com texto preto sobre fundo branco

Descrição gerada automaticamente

Resultado Preparação 1:

Tela de computador com texto preto sobre fundo branco

Descrição gerada automaticamente

Treinamento:

Texto

Descrição gerada automaticamente

Texto

Descrição gerada automaticamente

1. **Definição e descrição das bases teóricas dos métodos.**

O método a ser usado, inicialmente definido, é a Regressão Logística.

1. **Definição e descrição de como será calculada a acurácia**

A acurácia mede a eficiência de um modelo. No nosso escopo de trabalho, usaremos a acurácia para calcular quanto de sucesso obtivemos (em porcentagem) em prever a idade dos entrevistados. Também poderemos ver qual das medidas nos fornece uma melhor eficiência, se são as medidas fisiológicas ou bioquímicas.

Para isso, utilizaremos uma faixa de ±1. Por exemplo, caso a idade de uma pessoa seja 40 e o nosso modelo prever entre 39 e 41, podemos considerar como correto. Se tiver fora dessa faixa, será uma previsão incorreta.

Com isso, pegamos os valores preditos corretamente e dividimos pelo total de casos no modelo e obtemos a acurácia. Há algumas formas de fazermos isso no Python.

1. **Base de Dados/ Repositório do Github**

**Links das bases de dados:**

<https://archive.ics.uci.edu/dataset/887/national+health+and+nutrition+health+survey+2013-2014+(nhanes)+age+prediction+subset>

**Link para acesso ao GitHub:**

<https://github.com/pcmassonjr/ProjetoAplicado2>

1. **Cronograma de Atividades**

****