| COMPONENTE CURRICULAR: | Análise Exploratório de Dados |
| --- | --- |
| NOME COMPLETO DO ALUNO: | Anderson Alves, Gerson Soares Rodrigues, Samuel Bonfim da Silva, Andreia Domingos dos Santos |
| RA: | 10347602, 10423804, 10423569, 10288503 |
|  | |

**Tema Proposto: Predição multiclasse de risco de obesidade**

**Introdução**

A obesidade se tornou um desafio de saúde pública complexo e multifacetado que demonstra um crescimento contínuo nos últimos anos. Além de ser um risco bem estabelecido a uma série de problemas de saúde, entre eles as doenças cardiovasculares. Neste projeto, exploraremos a relação entre obesidade e doenças cardiovasculares, utilizando dados fornecidos pela empresa Kaggle para sustentar nossa análise inicial.

**Premissas e apresentação da empresa**

A empresa Kaggle é conhecida por suas competições entre o público de Data Science, com uma comunidade diversificada chegando a mais de 4 milhões de membros. Ela também é responsável por ajudar a comunidade, lançando alguns desafios com o intuito de praticar as habilidades sobre a análise de dados. Visando isso, aceitamos o desafio proposto pela Kaggle, da Tabular Playground Series, especificamente na quarta temporada, episódio 2, onde o nosso desafio será analisar dados fornecidos pela empresa sobre a “Multi-Class Prediction of Obesity Risk”. O objetivo é utilizar os vários fatores fornecidos, para prever o risco de obesidade em indivíduos, o qual está relacionado a doenças cardiovasculares.

**Objetivo e metas**

O objetivo principal do projeto é através dos dados obtidos, na plataforma Kaggle, adotar uma abordagem multifatorial para prever o risco de obesidade. Este risco é de suma importância, dada a sua associação direta com doenças cardiovasculares, as quais são grandes causas de morbidade e mortalidade ao redor do globo. Ao entender e antecipar os fatores que levam ao aparecimento e progressão da obesidade. Deste modo, teremos como metas, ao realizar a análise exploratória sobre os dados, adotar intervenções preventivas e terapêuticas mais assertivas. Essas intervenções abrangem a identificação precoce de padrões de risco e a customização de estratégias de intervenção, considerando os dados fornecidos pela plataforma. Assim, o foco do projeto abrange o desenvolvimento de modelos preditivos robustos, aspirando também a fornecer insights profundos que possam nortear políticas de saúde pública, práticas clínicas e esforços de promoção de saúde.

**Cronograma de Atividades**

**Etapa 1 - Kick-off do projeto**

1. **Montagem do Grupo**
   1. **Escolha dos participantes**
2. **Escolha da Temática**
   1. **Após uma pesquisa de artigos publicados na plataforma Kaggle, foi consenso de todos os participantes a escolha do tema a ser utilizado**
3. **Comunicação**
   1. **Foram feitas reuniões periódicas entre os participantes do grupo para a organização do assunto abordado da temática escolhida**
4. **Organização do Material**
   1. **Foram disponibilizados pela plataforma alguns dados os quais foram coletados (em formato CSV) que serão armazenados no Github do projeto e analisados nas próximas etapas;**
   2. **Os dados contêm valores numéricos e contínuos, então poderão ser utilizados para análise baseada em algoritmos de classificação, predição, segmentação e associação. Os dados estão disponíveis em formato CSV. Os conjuntos de dados que serão utilizados na análise são os seguintes:**

* **id - Identificação**
* **Gender - Gênero**
* **Age - Idade**
* **Height - Altura**
* **Weight - Peso**
* **family\_history\_with\_overweight - histórico familiar com sobrepeso**
* **FAVC - Frequência de consumo de alimentos calóricos**
* **FCVC - Frequência de consumo de Vegetais**
* **NCP - Número de refeições principais**
* **CAEC - Consumo de alimentos durantes das refeições**
* **SMOKE - Fumante ou não**
* **CH2O - Consumo diário de água**
* **SCC - Monitoramento de consumo calórico**
* **FAF - Frequência de Atividades Físicas**
* **TUE - Tempo Utilizado com Dispositivos Eletrônicos**
* **CALC - Consumo de Álcool**
* **MTRANS - Meio de transporte utilizado**
* **NObeyesdad - Nível de obesidade (Abaixo, Normal, Sobrepeso, Obesidade Grau I, Obesidade Grau II, Obesidade Grau IIII**

**Etapa 2 - Proposta do Produto Analítico e Análise Exploratória de Dados**

1. **Pensamento computacional em contextos organizacionais** 
   1. ***Definição do Problema:* Investigar a relação entre obesidade e doenças cardiovasculares diante do aumento da obesidade como um desafio de saúde pública.**
   2. ***Coleta de Dados:* Utilização de dados disponibilizados pela Kaggle através do desafio da Playground Series, temporada 4, episódio 2, focado na Previsão Multiclasse do Risco de Obesidade.**
   3. ***Análise:* Adoção de uma abordagem multifatorial para analisar os dados, visando prever o risco de obesidade e sua associação com doenças cardiovasculares.**
   4. ***Desenvolvimento do Modelo:* Criação de modelos preditivos baseados nos dados explorados para identificar padrões de risco e antecipar o surgimento da obesidade.**
   5. ***Intervenção:* Proposição de intervenções preventivas e terapêuticas personalizadas, fundamentadas na análise dos dados.**
   6. ***Impacto:* Fornecer insights para influenciar políticas de saúde pública, práticas clínicas e promover a saúde, enfrentando o desafio da obesidade e doenças relacionadas**

**Referência Bibliográfica**

* [**https://www.kaggle.com/competitions/playground-series-s4e2**](https://www.kaggle.com/competitions/playground-series-s4e2)
* **(Adicionar o Github do Projeto e inserir os dados que foram coletados no Kaggle)**