**Versão 1.0**

**MAR/2022**

Uma imagem contendo Texto

Descrição gerada automaticamente

**PLANO DE PROJETO**

**PREVENÇÃO DE diabetes**

**grupo: SPARK LABS**

# **PLANO DE PROJETO**

Este plano de projeto tem como objetivo descrever as informações necessárias para o desenvolvimento do projeto Prevenção de Diabetes.

## entendimento do negócio

Diabetes é uma doença crônica grave em que os indivíduos perdem a capacidade de regular efetivamente os níveis de glicose no sangue, podendo levar à redução da qualidade de vida e da expectativa de vida. Depois que diferentes alimentos são decompostos em açúcares durante a digestão, os açúcares são liberados na corrente sanguínea. Isso sinaliza o pâncreas para liberar insulina. A insulina ajuda a permitir que as células do corpo usem esses açúcares na corrente sanguínea para obter energia. A diabetes é geralmente caracterizada pelo corpo não produzir insulina suficiente ou ser incapaz de usar a insulina que é produzida de forma tão eficaz quanto necessária.

Complicações como doenças cardíacas, perda de visão, amputação de membros inferiores e doença renal estão associadas a níveis cronicamente altos de açúcar que permanecem na corrente sanguínea para pessoas com diabetes. Embora não haja cura para o diabetes, estratégias como perder peso, comer de forma saudável, ser ativo e receber tratamentos médicos podem mitigar os danos dessa doença em muitos pacientes. O diagnóstico precoce pode levar a mudanças no estilo de vida e tratamento mais eficaz, tornando os modelos preditivos de risco de diabetes ferramentas importantes para autoridades públicas e de saúde pública.

Embora existam diferentes tipos de diabetes, o diabetes tipo II é a forma mais comum e sua prevalência varia de acordo com idade, educação, renda, localização, raça e outros determinantes sociais da saúde. Grande parte do ônus da doença recai sobre aqueles de nível socioeconômico mais baixo também. A diabetes também representa um enorme fardo para a economia.

## escopo

O escopo do projeto será composto:

1. Elaboração de Plano de Projeto.
2. Criação de Data Pipeline para automatizar das etapas do projeto de dados (Ver necessidade de criar um Diagrama).
3. Criação de um Modelo Preditivo que dado um determinado conjunto de características o modelo consiga detectar se a pessoa está com alguma das doenças ou não e a sua probabilidade.
4. Criação de Dashboard para tomada de decisão.
5. Criação de Data App para consumo do modelo.
6. Elaboração de apresentação dos resultados.

## pessoas

A equipe do projeto será composta pelos seguintes profissionais:

| Perfil | Nome |
| --- | --- |
| Scrum Master: | Jacinto Xavier |
| Engenheiro de Dados: | Mateus Pereira |
| Cientista de Dados: | Douglas Lopes  Eduardo Moura  Jacinto Xavier  Matheus Fellipe |
| Analista de Dados: | Matheus Fellipe |

## métricas

Serão utilizadas as seguintes métricas:

* Acurácia
* Precision
* Recall
* F1-Score

## fases do projeto

**As fases do projeto estão aderentes a metodologia CRISP-DM:**

**Fase 1 – Entendimento do Negócio**

A primeira coisa a ser feita é entender de fato qual o problema/negócio a ser resolvido, buscando todos os detalhes sobre o impacto dele na empresa e quais os objetivos em relação ao trabalho.

**Fase 2 – Engenharia de Dados e Infraestrutura**

Nesta fase é definido a origem dos dados e infraestrutura que será necessária para o tratamento dos dados. Consiste em organizar e documentar todos os dados que se encontram disponíveis, dando início ao trabalho de mineração de dados, com objetivo de identificar quais são os dados importantes para a resolução do problema.

Nesse momento, o lado investigativo deve entrar em campo, para que os dados revelem problemas, soluções e tendências dos negócios.

**Fase 3 – Limpeza e Pré-Processamento**

Nesta fase são realizadas as limpezas de dados (eliminação de dados duplicados, nulos, outliers, etc) e pré-processamentos dos dados para eliminar os dados que não contribuirão para as análises das próximas fases.

**Fase 4 – Análise Exploratória**

Nesta etapa os dados já foram identificados, limpos e analisados, é hora de aplicar a parte técnica de análise deles. Serão preparadas os databases e definidos os formatos e questões técnicas da análise, sendo feita a escolha dos dados que serão trabalhados e de como eles serão cruzados para resolver o problema da empresa.

**Fase 5 – Modelagem**

Nesta fase são aplicadas as técnicas de Data Mining, com base nos objetivos identificados no primeiro momento. A mineração de dados pode ser associada a análises preditivas. Os dados minerados podem ser usados para alimentar algoritmos que preveem as tendências dos negócios.

**Fase 6 – Deploy e Apresentação**

Trata-se de um momento muito importante, pois é quando serão acompanhados os resultados em relação aos objetivos. Isso pode ser feito por meio de reuniões ou apresentações, onde os dados e insights são apresentados para os envolvidos nas tomadas de decisão.

Aqui, é a hora da verdade, onde tudo que foi obtido de conhecimento dos dados são entregues de forma a ser aplicada. A partir disso, podem ser mudados os processos dentro da organização e criados novos produtos — tudo com base em dados, garantindo, assim, o sucesso dos negócios.

Caso o resultado seja atingido, o modelo é colocado em produção (deploy).

## cronograma

Vide arquivo *Cronograma do Projeto de Prevenção de Diabetes.xlsx*

## ARQUITETURA

### DESENHO

Diagrama

Descrição gerada automaticamente

### FERRAMENTAS

Serão utilizadas as seguintes ferramentas:

* Anaconda (Jupyter Notebook)
* Word, Excel e Power Point
* GitHub Web e GitHub Desktop
* Databricks Community Edition
* Data Lake
* Visual Studio Code
* Power BI

### DADOS

**DADOS DE ENTRADA:**

* diabetes\_012\_health\_indicators\_BRFSS2015.csv

**DADOS TRANSFORMADOS:**

* analize\_exploratoria\_parquet.gzip (arquivo gerado após a limpeza dos dados e pré-processamento dos dados).
* modelagem\_parquet.gzip (arquivo gerado após a análise exploratória).

## COMUNICAÇÃO

Foram definidos 2 canais de comunicação:

Canal1 – Via WhatsApp (troca de mensagens diárias)

Canal2 – Sprints Semanais

* Planning (todas as segundas-feiras)
* Review (todas as sextas-feiras)