Slide 1

Sou Alessandra de assis barbosa e como trabalho de conclusão do curso de pós-graduação latu sensu em ciência de dados e big data da puc minas conduzi o estudo que teve por propósito avaliar a capacidade de previsão da expectativa de vida da população utilizando algoritmos de machine learning a partir de informações socioeconômicas para isso foram implementados modelos supervisionados de aprendizado de máquina.

Slide 2

A Expectativa de vida, também chamada de esperança de vida, é a expressão usada para indicar o número médio de anos que cada indivíduo provavelmente viverá caso sejam mantidas as mesmas condições vivenciadas no momento do nascimento. Em particular, a expectativa de vida ao nascer indica quantos anos em média um recém-nascido está destinado a viver.

Slide 3

O objetivo deste trabalho é realizar uma análise exploratória que permitirá de individuar os fatores correlacionados a redução ou aumento da expectativa de vida, identificar um modelo matemático que possa ser usado para a previsão da expectativa de vida, comparar a performance de vários modelos de machine learning e a realização de cenários que possam auxiliar na tomada de decisões de investimentos público e privado.

O objetivo maior deste projeto não é somente prever a expectativa de vida nos países mas identificar fatores diretamente vinculados**.**

Slide 4

Além disso esse trabalho também visa responder as seguintes hipótese:

●Os vários fatores de previsão escolhidos inicialmente realmente afetam a expectativa de vida?

● Quais são as variáveis de previsão que realmente afetam a expectativa de vida?

●Um país com expectativa de vida menor (<65) deve aumentar seus gastos com saúde para melhorar sua expectativa de vida média?

●Como as taxas de mortalidade de bebês e adultos afetam a expectativa de vida?

●A expectativa de vida tem correlação positiva ou negativa com hábitos alimentares, estilo de vida, exercícios, fumo, bebida alcoólica etc.

● Qual é o impacto da escolaridade na expectativa de vida dos humanos?

●A expectativa de vida tem uma relação positiva ou negativa com o consumo de álcool?

●Países densamente povoados ou altamente populosos tendem a ter menor expectativa de vida?

● Qual é o impacto da cobertura de imunização na expectativa de vida?

●Países mais poluídos apresentam uma expectativa de vida menor?

Slide 5

Na execuçao desse projeto foi utilizado uma metodologia ciclica para cada ciclo do projeto passamos por etapas que seram melhoradas a cada ciclo. As etapas aqui definidas sao: Coleta dos dados, Processamento/Tratamento de Dados, Análise e Exploração dos Dados, Criação de Modelos de Machine Learning, Interpretação dos Resultados, Apresentação dos Resultados

Slide 6

Coleta de dados

As informações utilizadas neste projeto são dados dos anos 2010 à 2015 e foram coletados em diferentes fontes: órgão mundial da saúde e das nações unidas, dados do IBGE, organização Our World in Data e API de geolocalização. São dados de 193 países que foram subdivididos em seis grandes áreas: fatores relacionados à economia, fatores ambientais, fatores demográficos, fatores de mortalidade, fatores relacionados à imunização e fatores relacionados à saúde

Slide 7

Processamento/Tratamento de Dados

Essa fase do projeto foi dividida em várias seções de forma a descrever com cura os dados obtidos.

Uma delas è a criação de features que è uma das técnicas de feature engineering que envolve derivar novas features das existentes. As features criadas no projeto sao:percentual de mulheres e emissão de gás por tamanho da população.

Outra seçao è analisar informaçoes do database como por exemplo dimesao do database e tipologia de dados.

Outra seçao faz uma analise descritiva dos dados , vemos estatisticas como kutosis , skewness nessa fase podemos ver algumas constataçoes. A análise descritiva dos dados ajuda a entender melhor a natureza dos dados, ver melhor os erros e entender melhor o processo e auxilia no tratamento dos dados e na escolha do algoritmo.

A ultima seçao avlia os dados omissos , nessa seçao foram usados alguns tipos de tratamentos dos dados faltantes: inserçao manual, nova coleta de dados , utilizo de algoritmos para prever dados, e eliminaçao de alguns dados

Slide 11

Análise e Exploração dos Dados

Nessa etapa realizamos a analise exploratoria de dados , foram criados varios tipos de graficos que mostram informaçoes e respondem algumas das hipoteses colocadas no inicio do projeto. Como por exemplo qual é o impacto da escolaridade na expectativa de vida dos humanos.

Slide 15

***Criação de Modelos de Machine Learning***