

# Inteligencia Artificial

Grupo - 55

David Vieira - 56550

Cláudio Silva - 54450

# Objetivo do Trabalho

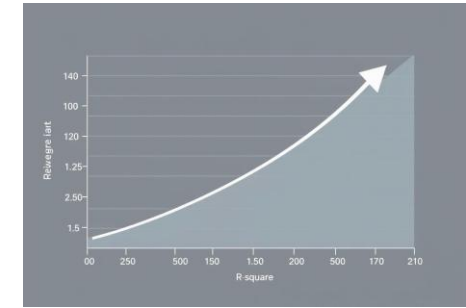
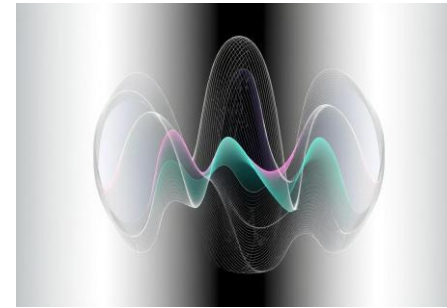
- Desenvolver um modelo de Machine Learning utilizando Support Vector Machines (SVM)
- **Objetivo principal:** Prever os custos futuros associados a novos clientes de uma companhia de seguros de saúde

O modelo deve ajudar a definir o valor do prémio de seguro de forma mais precisa e justa, estimando os custos para um conjunto de clientes fornecido (ficheiro `just_features.csv`).

Além disso, o projeto visa identificar as características que mais influenciam as despesas de saúde dos clientes.



# Metodologia



## Pré-processamento de Dados

Tratamento de valores ausentes, transformação de variáveis categóricas e normalização/padronização de variáveis numéricas.

## Divisão do Conjunto de Dados

Dados divididos em conjuntos de treino e teste (ex: 80% treino, 20% teste).

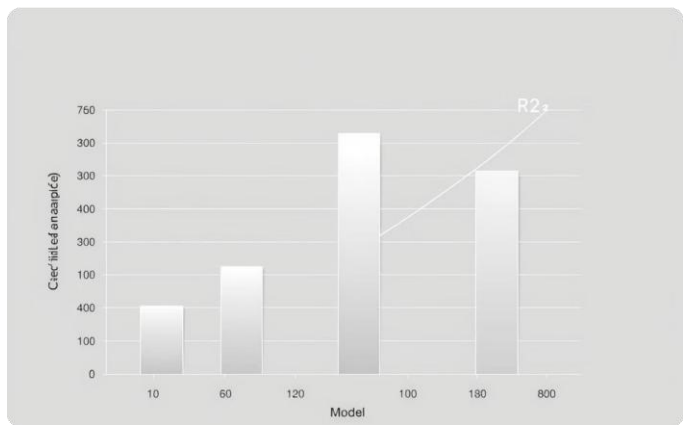
## Treinamento do Modelo

Algoritmo SVM para regressão, hiperparâmetros otimizados com GridSearchCV.

## Avaliação do Modelo

$R^2$  como métrica, desempenho avaliado no conjunto de teste.

# Resultados e Análise



## Desempenho do Modelo

O modelo atingiu um alto coeficiente de determinação ( $R^2$ ), demonstrando precisão na previsão de custos.

		Curtorays	Pasty (Alforica	Pasty Alforica	Jamph (Alforica	Pastl	Usofofoesica
1	Piebuca	SuteCD	383,0000	105,130000	395,40000	291,80000	741,60000
1	Famionica	SuteCD	361,6000	1315,50000	395,40000	996,60000	552,60000
2	Eulicoica	SateCD	251,0000	305,300000	602,40000	385,90000	797,90000
3	Faulraice	SateCD	241,9000	461,00000	597,70000	966,60000	782,40000
4	Fatlocica	SuteCD	331,0000	166,210000	167,40000	223,80000	767,50000
5	Posiraice	SateCD	261,0000	122,50000	267,40000	256,30000	753,60000
4	Fabilarica	SuteCD	233,7000	106,100000	225,20000	395,40000	755,90000
10	Prolinica	SuteCD	261,6000	327,210000	299,40000	396,80000	185,30000
13	Implinica	SateCD	762,6000	212,300000	182,50000	281,30000	765,40000
18	Wierforica	CuteCD	282,9000	199,200000	226,40000	331,90000	765,30000

## Estimativas e Prévisões

Os custos previstos foram exportados para o arquivo grupo\_custos\_estimados.csv.



## Análise das Características

### Importantes

Análise de remoção incremental de características identificou os principais fatores que influenciam os custos de saúde.

# Conclusões e Trabalhos Futuros

1

## Conclusões

O modelo SVM demonstrou capacidade de prever os custos futuros com uma precisão aceitável, medido pelo  $R^2$  obtido.

As características com maior impacto nos custos foram identificadas (ex.: IMC, hábito de fumar, classe etária, etc.).

O modelo pode ser utilizado como ferramenta auxiliar na definição de prémios de seguros de saúde de forma mais justa e precisa.

2

## Limitações

O modelo pode apresentar algumas limitações, como dependência da qualidade dos dados ou possível necessidade de otimização dos hiperparâmetros.

O modelo SVM pode ser computacionalmente intensivo para grandes volumes de dados.

# Agradecimentos e Referências

Agradecemos a oportunidade de desenvolver este projeto, que nos permitiu aplicar os conhecimentos adquiridos em Inteligência Artificial e consolidar a prática com algoritmos de Machine Learning

## Referências

- Scikit-learn: <https://scikit-learn.org>
- Documentação Pandas: <https://pandas.pydata.org>