

Empresa TechSolutions

Nome da Empresa: TechSolutions

Ramo da Empresa: Tecnologia da Informação

Link Github: [<https://github.com/maxstonne/TechSolutions>]

Equipe Responsável:

Mateus Amarante Amaral - RA: 2223104622

Curso: Analise e Desenvolvimento de Sistemas

Turma: A

Semestre: 5

Ano: 2024

Índice

Escopo do Projeto:	3
Serviços Oferecidos:	3
Estruturação Interna da Empresa	4
3.1 - Aprendizado de Máquina	4
3.1.1 - Regressão Linear	4
3.1.2 - Classificação com SVM	5
3.2 - Ciência de Dados	6
3.2.1 - Análise Estatística Descritiva	6
3.2.2 - Análise de Correlação e Visualização.....	7
3.3 - Modelagem de Dados	8
3.3.1 - Banco de Dados Relacional	8
3.3.2 - Banco de Dados Dimensional	9
3.4 - Redes de Computadores	10
3.4.1 - Cálculo da Máscara de Sub-rede	10
3.4.2 - Simulação de Ping Simples	11
3.5 - Segurança da Informação	12
3.5.1 - Criptografia e Descriptografar de Texto	12
3.5.2 - Cálculo do Hash de um Arquivo.....	13

Escopo do Projeto:

O presente documento detalha o escopo do projeto proposto pela empresa **TechSolutions**. O projeto tem como objetivo desenvolver soluções inovadoras para a segurança da informação, visando proteger dados sensíveis e garantir a integridade das informações em diversas plataformas. Este projeto abrange criptografia de dados, verificação de integridade de arquivos, e implementação de técnicas de aprendizado de máquina e ciência de dados para aprimorar a segurança.

Serviços Oferecidos:

A empresa TechSolutions oferecerá os seguintes serviços:

- **Desenvolvimento de Algoritmos de Aprendizado de Máquina:**
 - Aplicação de modelos preditivos e análise de dados para otimização de processos empresariais.
- **Consultoria em Ciência de Dados:**
 - Implementação de soluções de análise de dados e estatísticas avançadas para tomada de decisão.
- **Modelagem de Banco de Dados:**
 - Criação e gestão de bases de dados eficientes e seguras.
- **Infraestrutura de Redes de Computadores:**
 - Planejamento, instalação e manutenção de redes de computadores.
- **Segurança da Informação:**
 - Desenvolvimento e implementação de políticas de segurança e proteção de dados.

Estruturação Interna da Empresa

3.1 - Aprendizado de Máquina

3.1.1 - Regressão Linear

Conjunto de Dados: O conjunto de dados utilizado é o Boston Housing, que contém informações sobre preços de casas em Boston. As características incluem, entre outras, a taxa de criminalidade, o índice de qualidade das escolas, e a proximidade a pontos de interesse.

Resultados do Modelo:

- **Erro Quadrático Médio (Mean Squared Error - MSE):** 24.29
 - **Interpretação:** O MSE mede a média dos quadrados dos erros, isto é, a diferença média entre os valores previstos e os valores reais. Um MSE de 24.29 indica que, em média, as previsões do modelo diferem dos valores reais dos preços das casas em 24.29 unidades de preço quadrático.
- **Coeficiente de Determinação (R-squared):** 0.68
 - **Interpretação:** O R-squared é uma medida que indica a proporção da variabilidade nos dados dependentes que é explicada pelo modelo. Um valor de 0.68 significa que 68% da variação nos preços das casas é explicada pelas características consideradas no modelo.

Análise

O modelo de regressão linear mostra um bom ajuste aos dados com um R-squared de 0.68, o que indica que o modelo consegue explicar uma boa parte da variabilidade dos preços das casas em Boston com base nas características fornecidas. O MSE de 24.29 sugere que há espaço para melhorar a precisão do modelo, talvez considerando outras características ou utilizando métodos de regressão mais complexos.

3.1.2 - Classificação com SVM

Conjunto de Dados: O conjunto de dados utilizado é o Iris, que contém informações sobre três espécies de flores (Setosa, Versicolor e Virginica). As características incluem o comprimento e a largura das sépalas e pétalas.

- **Acurácia (Accuracy):** 1.0
 - **Interpretação:** A acurácia mede a proporção de previsões corretas feitas pelo modelo. Uma acurácia de 1.0 (100%) indica que o modelo classificou corretamente todas as instâncias no conjunto de teste.
- **Relatório de Classificação (Classification Report):**
 - **Precision, Recall, F1-Score:**
 - Todas as três classes (Setosa, Versicolor, Virginica) têm uma precisão, recall e F1-score de 1.0.

Interpretação:

- **Precision:** Proporção de verdadeiros positivos sobre o total de positivos preditos. Uma precisão de 1.0 para todas as classes indica que todas as flores classificadas como pertencentes a uma determinada espécie realmente pertencem a essa espécie.
- **Recall:** Proporção de verdadeiros positivos sobre o total de positivos reais. Um recall de 1.0 indica que o modelo identificou corretamente todas as instâncias de cada espécie.
- **F1-Score:** Média harmônica de precision e recall. Um F1-score de 1.0 indica um equilíbrio perfeito entre precisão e recall.
- **Matriz de Confusão (Confusion Matrix):** A matriz de confusão mostra que todas as previsões do modelo estão corretas, sem nenhuma instância sendo classificada erroneamente.

Análise

O modelo de SVM apresenta um desempenho perfeito no conjunto de dados Iris, com uma acurácia de 100% e todas as métricas de desempenho (precision, recall, F1-score) também sendo perfeitas. Isso sugere que o conjunto de dados é bem separado nas suas características e que o modelo SVM é altamente eficaz para esta tarefa de classificação.

3.2 - Ciência de Dados

3.2.1 - Análise Estatística Descritiva

Conjunto de Dados: O conjunto de dados utilizado é o Iris, que contém informações sobre três espécies de flores (Setosa, Versicolor e Virginica). As características incluem o comprimento e a largura das sépalas e pétalas.

Estatísticas Descritivas

- **Média (Mean):**
 - sepal length (cm): 5.843
 - sepal width (cm): 3.054
 - petal length (cm): 3.759
 - petal width (cm): 1.199
- **Mediana (Median):**
 - sepal length (cm): 5.80
 - sepal width (cm): 3.00
 - petal length (cm): 4.35
 - petal width (cm): 1.30
- **Moda (Mode):**
 - sepal length (cm): 5.0
 - sepal width (cm): 3.0
 - petal length (cm): 1.5
 - petal width (cm): 0.2
- **Desvio Padrão (Standard Deviation):**
 - sepal length (cm): 0.828
 - sepal width (cm): 0.434
 - petal length (cm): 1.764
 - petal width (cm): 0.763

Visualização da Distribuição

- **Histograma com KDE:** Mostra a distribuição das características sepal length, sepal width, petal length, e petal width. As distribuições ajudam a visualizar como os valores se espalham e se há uma tendência central.
- **Boxplot:** Apresenta os valores mínimos, primeiros quartis, medianas, terceiros quartis e máximos das características. Também destaca os outliers, que são pontos fora do intervalo interquartil.

Análise

Os dados mostram que há uma variabilidade considerável nas características das flores, com petal length e petal width tendo maiores desvios padrão, indicando maior dispersão. As medianas e modas fornecem uma ideia de valores centrais, enquanto os boxplots ajudam a identificar possíveis outliers.

3.2.2 - Análise de Correlação e Visualização

Visualização da Correlação

- **Mapa de Calor:** O mapa de calor da matriz de correlação mostra a intensidade das relações entre pares de características. Valores mais próximos de 1 ou -1 indicam uma correlação forte, positiva ou negativa, respectivamente. Por exemplo, petal length e petal width têm uma correlação muito forte (0.962865).
- **Pairplot:** Os scatterplots mostram a relação entre todas as características, diferenciando as espécies de flores com cores diferentes. Isso ajuda a identificar visualmente as correlações e as separações entre as espécies.

Análise

A análise de correlação revela que há fortes correlações positivas entre sepal length e petal length, sepal length e petal width, e petal length e petal width. A correlação negativa entre sepal width e as outras características sugere uma relação inversa mais fraca. Os scatterplots no pairplot confirmam visualmente essas correlações e ajudam a entender a distribuição das diferentes espécies de flores no espaço das características.

3.3 - Modelagem de Dados

3.3.1 - Banco de Dados Relacional

Para o banco de dados relacional, criamos um sistema para gerenciar uma biblioteca, incluindo informações sobre autores e livros.

Estrutura das Tabelas

1. Tabela Autor:

- **id:** Identificador único do autor.
- **nome:** Nome do autor.

2. Tabela Livro:

- **id:** Identificador único do livro.
- **título:** Título do livro.
- **genero:** Gênero do livro.
- **autor_id:** Referência ao autor do livro, conectando-se com a tabela Autor.

Dados Obtidos

- **Autores:**
 - Autor 1: J.K. Rowling
 - Autor 2: George R.R. Martin
- **Livros:**
 - Livro 1: "Harry Potter e a Pedra Filosofal", Gênero: Fantasia, Autor: J.K. Rowling
 - Livro 2: "A Game of Thrones", Gênero: Fantasia, Autor: George R.R. Martin

Banco de Dados Relacional:

- Estrutura: Tabelas de autores e livros com uma relação de um para muitos.
- Uso: Gerenciamento de informações literárias e suas relações.

Análise

Os dados relacionam autores e livros, onde cada livro é associado a um autor específico. Isso permite gerenciar quais livros pertencem a quais autores de maneira organizada e estruturada.

3.3.2 - Banco de Dados Dimensional

Para o banco de dados dimensional, criamos um sistema para gerenciar vendas, incluindo informações sobre produtos, clientes e datas.

Estrutura das Tabelas

1. Tabela Produto:

- **id:** Identificador único do produto.
- **nome:** Nome do produto.
- **categoria:** Categoria do produto.

2. Tabela Cliente:

- **id:** Identificador único do cliente.
- **nome:** Nome do cliente.
- **cidade:** Cidade do cliente.

3. Tabela Tempo:

- **id:** Identificador único da data.
- **data:** Data completa (ano-mês-dia).
- **ano:** Ano da data.
- **mes:** Mês da data.
- **dia:** Dia da data.

4. Tabela Vendas:

- **id:** Identificador único da venda.
- **produto_id:** Referência ao produto vendido, conectando-se com a tabela Produto.
- **cliente_id:** Referência ao cliente que fez a compra, conectando-se com a tabela Cliente.
- **tempo_id:** Referência ao tempo da venda, conectando-se com a tabela Tempo.
- **quantidade:** Quantidade de produtos vendidos.
- **valor_total:** Valor total da venda.

Dados Obtidos

- **Produtos:**
 - Produto 1: Laptop, Categoria: Eletrônicos
 - Produto 2: Cadeira, Categoria: Móveis
- **Clientes:**
 - Cliente 1: João Silva, Cidade: São Paulo
 - Cliente 2: Maria Oliveira, Cidade: Rio de Janeiro
- **Tempo:**
 - Tempo 1: Data: 2024-01-01, Ano: 2024, Mês: janeiro, Dia: 1
 - Tempo 2: Data: 2024-01-02, Ano: 2024, Mês: janeiro, Dia: 2
- **Vendas:**
 - Venda 1: Produto: Laptop, Cliente: João Silva, Data: 2024-01-01, Quantidade: 2, Valor Total: 3000.0
 - Venda 2: Produto: Cadeira, Cliente: Maria Oliveira, Data: 2024-01-02, Quantidade: 1, Valor Total: 500.0

Banco de Dados Dimensional:

- Estrutura: Tabelas de produtos, clientes, tempo e vendas, com tabelas de fatos e dimensões.
- Uso: Análise de vendas, permitindo uma visão detalhada das transações por produto, cliente e tempo.

Análise

Os dados relacionam vendas com produtos, clientes e datas, permitindo análises detalhadas sobre o desempenho das vendas, perfil dos clientes e distribuição temporal das transações. Esse modelo é adequado para análises complexas e relatórios gerenciais.

3.4 - Redes de Computadores

3.4.1 - Cálculo da Máscara de Sub-rede

Este algoritmo calcula a máscara de sub-rede correspondente a um endereço IP e um número de bits da máscara de sub-rede fornecidos pelo usuário.

Dados Obtidos

1. **Endereço IP:** O endereço IP fornecido pelo usuário (por exemplo, 192.168.1.1).
2. **Bits da Máscara:** O número de bits da máscara de sub-rede fornecido pelo usuário (por exemplo, 24).
3. **Máscara de Sub-rede:** A máscara de sub-rede calculada com base nos dados de entrada (por exemplo, 255.255.255.0).

Análise dos Dados

- **Endereço IP:** Este é o endereço IP que o usuário deseja utilizar para calcular a máscara de sub-rede.
- **Bits da Máscara:** Este número indica a quantidade de bits utilizados na máscara de sub-rede, determinando a parte da rede e a parte do host no endereço IP.
- **Máscara de Sub-rede:** Este é o resultado do cálculo, que mostra qual parte do endereço IP é a rede e qual parte é o host.

Exemplo:

- **Endereço IP:** 192.168.1.1
- **Bits da Máscara:** 24
- **Máscara de Sub-rede:** 255.255.255.0

3.4.2 - Simulação de Ping Simples

Este algoritmo verifica a conectividade com um host fornecido pelo usuário, simulando o comando de ping.

Dados Obtidos

1. **Host:** O nome de host ou endereço IP do destino a ser pingado (por exemplo, google.com).
2. **Resultado do Ping:** A resposta do comando de ping, indicando se o host está acessível ou não.

Análise dos Dados

- **Host:** Este é o nome do host ou o endereço IP que o usuário deseja verificar a conectividade.
- **Resultado do Ping:** Este resultado indica se o host está acessível a partir do sistema onde o algoritmo foi executado. O resultado pode ser:
 - **Acessível:** Significa que o host respondeu ao comando de ping.
 - **Não acessível:** Significa que o host não respondeu ao comando de ping.

Exemplo:

- **Host:** google.com
- **Resultado:** google.com está acessível ou google.com não está acessível

Resumo dos Dados Obtidos

1. Cálculo da Máscara de Sub-rede:

- **Entrada:** Endereço IP (192.168.1.1) e bits da máscara (24).
- **Saída:** Máscara de sub-rede calculada (255.255.255.0).

2. Simulação de Ping Simples:

- **Entrada:** Nome de host ou endereço IP (google.com).
- **Saída:** Resultado do ping indicando a acessibilidade do host (google.com está acessível ou google.com não está acessível).

Esses dados demonstram a funcionalidade e a eficácia dos algoritmos na realização de suas respectivas tarefas de cálculo de máscara de sub-rede e verificação de conectividade de rede.

3.5 - Segurança da Informação

3.5.1 - Criptografia e Descriptografar de Texto

Este algoritmo realiza a criptografia e a descriptografia de um texto utilizando o método de criptografia simétrica Fernet da biblioteca cryptography.

Dados Obtidos

1. Texto Original:

- **Entrada:** Texto em claro que o usuário deseja criptografar (por exemplo, "Segurança da Informação é crucial.").
- **Tipo:** String.

2. Chave de Criptografia:

- **Entrada:** Chave gerada pelo algoritmo para criptografar e descriptografar o texto.
- **Tipo:** Bytes.

3. Texto Criptografado:

- **Saída:** Texto criptografado, convertido em formato binário e ilegível sem a chave correta.
- **Tipo:** Bytes.

4. Texto Descriptografado:

- **Saída:** Texto original recuperado a partir do texto criptografado utilizando a chave correta.
- **Tipo:** String.

Exemplo

- **Texto Original:** "Segurança da Informação é crucial."
- **Chave de Criptografia:**
b'F3Vu7S45VwUy7EN8s8k23UbLXyltNi3oFxxsS31bm9Y='
- **Texto Criptografado:** b'gAAAAABgFwQWb...k1s=' (formato binário)
- **Texto Descriptografado:** "Segurança da Informação é crucial."

Análise dos Dados

- **Segurança:** A chave de criptografia deve ser mantida em segurança, pois é essencial para a recuperação do texto original.
- **Confidencialidade:** O texto criptografado garante que os dados sejam protegidos contra acesso não autorizado.
- **Integridade:** O texto descriptografado deve corresponder exatamente ao texto original, garantindo que não houve alterações.

3.5.2 - Cálculo do Hash de um Arquivo

Este algoritmo calcula o hash SHA-256 de um arquivo para garantir a integridade dos dados.

Dados Obtidos

1. Caminho do Arquivo:

- **Entrada:** Caminho para o arquivo cujo hash será calculado (por exemplo, "exemplo.txt").
- **Tipo:** String.

2. Hash SHA-256:

- **Saída:** Hash hexadecimal do arquivo, que é uma representação única do conteúdo do arquivo.
- **Tipo:** String.

Exemplo

- **Caminho do Arquivo:** "exemplo.txt"
- **Hash SHA-256:** 9e107d9d372bb6826bd81d3542a419d6 (exemplo)

Análise dos Dados

- **Integridade:** O hash SHA-256 garante que o arquivo não foi alterado. Se o conteúdo do arquivo mudar, o hash também mudará.
- **Autenticidade:** O hash pode ser usado para verificar se o arquivo recebido é autêntico e não foi adulterado.
- **Segurança:** O uso de SHA-256 proporciona um alto nível de segurança, pois é resistente a colisões, tornando difícil gerar dois arquivos diferentes com o mesmo hash.

Resumo dos Dados Obtidos

1. Criptografia e Descriptografia de Texto:

- **Entrada:** Texto original e chave de criptografia.
- **Saída:** Texto criptografado e texto descriptografado.
- **Aplicabilidade:** Garante a confidencialidade dos dados, protegendo o texto de leitura não autorizada.

2. Cálculo do Hash de um Arquivo:

- **Entrada:** Caminho para o arquivo.
- **Saída:** Hash SHA-256 do arquivo.
- **Aplicabilidade:** Garante a integridade dos dados, verificando se o arquivo não foi alterado.

Esses dados demonstram a eficácia dos algoritmos na realização de suas respectivas tarefas de criptografia e verificação de integridade, fundamentais para a segurança da informação.