

FRAMEWORK DE GERAÇÃO DE DATASET BASEADO EM OCR PARA CÁLCULO DE CONSUMO MENSAL DE ENERGIA ELÉTRICA

Christiano Talhate (IFES-Instituto Federal do Espírito Santo) talhate@hotmail.com

Resumo

Esse trabalho apresenta uma abordagem inovadora combinando técnicas de processamento de imagens e aprendizado de máquina para automatizar a coleta e análise de dados necessários para prever o consumo de energia elétrica em um determinado período. Esse tipo de framework utiliza algoritmos de OCR para extrair informações relevantes dos dados presentes em faturas de energia elétrica, como datas, leituras do medidor, tarifas e outros detalhes relacionados ao consumo. Esses dados extraídos são utilizados para construir um dataset estruturado, contendo as variáveis necessárias para o algoritmo de regressão. Ao utilizar o OCR para a extração de dados, o processo de coleta de informações de consumo mensal de energia elétrica se torna mais ágil e eficiente, eliminando a necessidade de inserção manual dos dados. Além disso, o framework pode ser treinado para tratar diferentes formatos e layouts de faturas, aumentando a flexibilidade e adaptabilidade do sistema. Com o dataset estruturado gerado pelo framework, o algoritmo de regressão pode ser aplicado para realizar a análise estatística dos dados e identificar padrões de consumo de energia ao longo do tempo permitindo prever com maior precisão o consumo mensal em um determinado cliente, fornecendo informações para o planejamento energético e a gestão eficiente dos recursos.

Palavras-Chaves: *Framework, OCR, DatasetEnergia, PrevisãoConsumoEnergia*

1. Introdução

A previsão do consumo de energia elétrica tornou-se um assunto de extrema importância tanto para empresas do setor energético quanto para os consumidores. Trata-se de um desafio fundamental que está intimamente ligado ao comportamento da demanda de energia necessária para o funcionamento adequado de uma organização, levando em consideração a quantidade de energia contratada. No entanto, atualmente não existe uma metodologia adequada para dimensionar o consumo corporativo de forma eficiente.

Com o avanço tecnológico e a disponibilidade crescente de dados, a análise do consumo de energia elétrica tem se beneficiado significativamente. A capacidade de prever com precisão o consumo mensal de energia elétrica tornou-se crucial para setores como planejamento energético, gerenciamento de demanda e faturamento. Essas previsões precisas possibilitam um melhor planejamento do uso dos recursos energéticos, evitando desperdícios e permitindo uma alocação mais eficiente dos recursos disponíveis.

Nesse contexto, identificamos a necessidade de um framework de geração de dataset baseado em OCR para alimentar algoritmos de regressão que auxiliem na previsão do consumo mensal de energia elétrica. A utilização de Optical Character Recognition (OCR) oferece uma solução promissora para extrair informações valiosas a partir das faturas de energia elétrica, permitindo a criação de conjuntos de dados adequados para treinar os modelos de regressão. Dessa forma, o presente trabalho propõe uma solução baseada em inteligência artificial que recupera dados de consumo de uma série histórica de contas de energia elétrica, dados recuperados de forma automática através da tecnologia de OCR e que fornece como resultado um dataset consolidado para pronto uso por algoritmos de regressão para maior previsibilidade de contratação de energia pela organização.

Essa solução tem como escopo desenvolver um framework que fornece um dataset onde busca-se resolver o desafio de obter dados precisos e confiáveis, que serão utilizados para treinar os algoritmos de regressão. Com isso, espera-se obter resultados mais acurados nas previsões de consumo mensal de energia elétrica. Esses resultados terão impactos significativos no planejamento energético das empresas, no gerenciamento da demanda e na melhoria do processo de faturamento.

Como escopo desse trabalho, teremos os seguintes pontos:

2. Visão geral do problema

2.1- O desafio de prever o consumo mensal de energia elétrica

2.2- A importância da obtenção de dados precisos para os algoritmos de regressão

3. Fundamentos de OCR

3.1- Princípios básicos do Optical Character Recognition (OCR)

3.2- Algoritmos utilizados na extração de informações

4. Desenvolvimento do framework



4.1- Processo de desenvolvimento do framework

4.2- Arquitetura de software

4.3- Arquitetura de solução em nuvem

4.4- Componentes

5. Indicação de algoritmos de regressão

5.1- Descrição de algoritmos de regressão

6. Resultados e aplicações

6.1- Principais aplicações práticas

6.2- Análise sobre a qualidade e confiabilidade das informações extraídas das faturas de energia elétrica

Esperamos que o framework de geração de dataset baseado em OCR seja capaz de facilitar a obtenção de dados para treinamento de modelos de regressão de consumo mensal de energia elétrica. Isso permitirá melhorar a precisão das previsões, contribuindo para uma gestão mais eficiente dos recursos energéticos e tomada de decisões mais informadas.

2-Desenvolvimento

A apresentação do processo de desenvolvimento do framework de geração de dataset é uma oportunidade importante para compartilhar informações sobre como o framework foi concebido, projetado e implementado.

2.1- Visão geral do problema

Essa sessão tem como objetivo permitir uma compreensão mais completa e abrangente da situação, facilitando a identificação de soluções eficazes, permitindo delinear estratégias e abordagens adequadas de forma relevante para apresentar uma solução eficaz, levando em consideração todos os seus aspectos relevantes e perspectivas futuras.

2.2- O desafio de prever o consumo mensal de energia elétrica

A previsão do consumo mensal de energia elétrica é um desafio importante e relevante no setor de energia. A capacidade de prever com precisão o consumo de energia elétrica tem várias aplicações práticas, tais como o planejamento da capacidade de geração, a otimização da distribuição de energia, a implementação de políticas de eficiência energética e a gestão da demanda.

Uma abordagem comum para prever o consumo mensal de energia elétrica é utilizar algoritmos de regressão, que procuram identificar padrões e relações entre os dados nos históricos de consumo existente nas contas de energia permitindo que o modelo possa ser treinado para aprender os padrões existentes e, em seguida, utilizar esse aprendizado para realizar previsões futuras.

No entanto, o desafio da previsão do consumo mensal de energia elétrica vai além da escolha do algoritmo de regressão. É necessário considerar cuidadosamente as variáveis relevantes para o problema, a qualidade dos dados disponíveis e o comportamento do consumidor.

A avaliação dos modelos de previsão é outro aspecto crítico. Métodos como validação cruzada ou divisão dos dados em conjuntos de treinamento e teste podem ser aplicados para medir a acurácia das previsões. O desempenho do modelo pode ser avaliado utilizando métricas como o erro médio absoluto (MAE) ou o erro médio quadrático (RMSE), que indicam o quão bem as previsões se aproximam dos valores reais.

Em resumo, a previsão do consumo mensal de energia elétrica é um desafio complexo que envolve a utilização de técnicas de regressão e análise de dados. Essa previsão precisa e confiável pode trazer benefícios significativos para o setor de energia, permitindo uma melhor gestão e planejamento da oferta e demanda de energia elétrica.

2.3- A importância da obtenção de dados precisos para os algoritmos de regressão

A obtenção de dados precisos é fundamental para alimentar os algoritmos de regressão de forma eficaz e obter previsões acuradas. A qualidade dos dados utilizados no treinamento dos algoritmos tem um impacto direto na qualidade das previsões geradas. Na Tabela 1, existem algumas razões pelas quais a obtenção de dados precisos é importante:

Tabela 1 – Importância para obtenção de dados precisos

RAZÃO	PORQUE É IMPORTANTE?
Acurácia das previsões	Os algoritmos de regressão baseiam-se nos dados históricos para identificar padrões e relações entre as variáveis. Se os dados estiverem imprecisos, com erros ou inconsistências, o modelo de regressão pode aprender padrões incorretos ou produzir previsões enviesadas. Dados imprecisos podem levar a resultados distorcidos e não confiáveis.
Tomada de decisões informadas	As previsões baseadas em algoritmos de regressão são frequentemente usadas como base para a tomada de decisões em várias áreas, como planejamento energético, gestão de demanda e otimização de recursos. Se os dados alimentados nos algoritmos não forem precisos, as decisões tomadas com base nessas previsões podem ser inadequadas e resultar em ações ineficientes ou ineficazes.
Identificação de fatores influentes	A regressão é usada para identificar quais variáveis têm um impacto significativo no resultado previsto. Se os dados não forem precisos, a identificação desses fatores influentes pode ser comprometida. Dados imprecisos podem levar a conclusões errôneas sobre quais variáveis estão realmente relacionadas ao fenômeno em estudo.
Melhoria contínua do modelo	A obtenção de dados precisos também permite uma melhor avaliação do desempenho do modelo de regressão. Com dados confiáveis, é possível comparar as previsões geradas pelo modelo com os valores reais observados e identificar oportunidades de melhoria. Dados imprecisos podem dificultar a avaliação adequada do modelo e prejudicar os esforços de otimização e refinamento.

Fonte: Elaborado pelo autor.

Para garantir a obtenção de dados precisos, é importante adotar boas práticas de coleta, limpeza e verificação dos dados. Isso envolve a utilização de fontes confiáveis, a validação dos dados em relação a critérios pré-estabelecidos, a detecção e correção de erros ou inconsistências e a manutenção de um processo contínuo de monitoramento da qualidade dos dados.

Dessa forma, a obtenção de dados precisos é essencial para alimentar os algoritmos de regressão de forma adequada, pois dados imprecisos podem comprometer a qualidade das previsões, influenciar negativamente a tomada de decisões e levar a conclusões errôneas, portanto, garantir a precisão dos dados é um passo fundamental para obter resultados confiáveis e informados a partir dos algoritmos de regressão.

3. Fundamentos de OCR

A tecnologia de Reconhecimento Óptico de Caracteres (OCR) tem revolucionado a forma como lidamos com informações impressas. Imagine poder transformar documentos físicos em texto digital editável, pesquisável e processável em apenas alguns segundos. Isso é possível graças aos fundamentos da OCR, que combinam técnicas avançadas de processamento de imagem e aprendizado de máquina

3.1- Princípios básicos do Optical Character Recognition (OCR)

O Optical Character Recognition (OCR), ou Reconhecimento Óptico de Caracteres, é uma tecnologia que permite a extração automática de texto e informações de imagens ou documentos digitalizados. O OCR é amplamente utilizado em diversas aplicações, como processamento de documentos, automação de escritórios, digitalização de livros, reconhecimento de placas de veículos e muito mais.

Os princípios básicos do OCR envolvem um conjunto de etapas que permitem a conversão de uma imagem contendo texto em texto editável ou estruturado. Essas etapas podem variar de acordo com as implementações e as tecnologias utilizadas, e geralmente envolvem os passos apresentados na Figura 1.

Figura 1-Etapas de extração de dados via OCR



Fonte: Elaborado pelo autor

É importante mencionar que o desempenho do OCR pode variar dependendo de vários fatores, como qualidade da imagem, resolução, iluminação, tipo e estilo da fonte, linguagem do texto e complexidade do layout do documento. Além disso, o OCR pode enfrentar desafios adicionais em casos de texto manuscrito, caracteres ilegíveis, documentos danificados ou em condições adversas de captura.

O OCR é uma tecnologia poderosa que permite a extração automática de texto e informações a partir de imagens ou documentos digitalizados, onde seus princípios básicos envolvem o pré-processamento da imagem, a segmentação de caracteres, o reconhecimento de caracteres e o pós-processamento para obter uma saída de texto reconhecido.

3.2- Algoritmos utilizados na extração de informações

A extração de informações de documentos é uma tarefa fundamental em diversas áreas, incluindo a recuperação de informações e mineração de dados. Essa atividade consiste em identificar e capturar informações relevantes contidas em documentos, como textos, imagens, tabelas ou qualquer outro tipo de conteúdo estruturado ou não estruturado. Existem várias técnicas e algoritmos que são frequentemente utilizados na extração de informações de documentos, tais como:

Processamento de Linguagem Natural (PLN): Conforme Weizenbaum (1966), o processamento de linguagem natural é a ponte que permite que humanos e máquinas se comuniquem de forma eficaz, abrindo um mundo de possibilidades na interação homem-máquina. O PLN é uma área de estudo que se dedica a desenvolver técnicas e algoritmos para processar e compreender a linguagem humana. Na extração de informações de documentos, o PLN é amplamente utilizado para realizar tarefas como segmentação de sentenças, tokenização, análise sintática, análise semântica e identificação de entidades nomeadas. Essas técnicas são essenciais para identificar e extrair informações específicas, como nomes de pessoas, datas, locais e outras entidades relevantes.

Reconhecimento Óptico de Caracteres (OCR): O OCR é a ponte que permite aos documentos em papel cruzarem para o mundo digital, tornando-se acessíveis, pesquisáveis e aproveitáveis. O OCR é uma tecnologia que permite converter imagens de textos em caracteres editáveis. É amplamente utilizado para extrair informações de documentos físicos, como livros, revistas, contratos e faturas. Algoritmos de OCR analisam as imagens, identificam os caracteres

presentes nelas e os convertem em texto legível. Essa técnica é particularmente útil quando se deseja extrair informações de documentos escaneados ou digitalizados.

Algoritmos de Aprendizado de Máquina: Os algoritmos de aprendizado de máquina desempenham um papel fundamental na extração de informações de documentos. Esses algoritmos podem ser treinados para reconhecer padrões e estruturas em textos, imagens ou outros tipos de conteúdo. Por exemplo, é possível treinar um modelo de aprendizado de máquina para identificar campos específicos em um formulário, como nome, endereço e número de telefone. Essa abordagem é amplamente utilizada em processos de automação de documentos, nos quais se deseja extrair informações consistentes de um grande número de arquivos.

Processamento de Imagens: Para documentos que contêm informações em formato de imagem, como gráficos, diagramas ou logotipos, técnicas de processamento de imagens podem ser aplicadas. Essas técnicas envolvem a análise de características visuais, como formas, cores e texturas, para extrair informações relevantes. Algoritmos de segmentação, detecção de bordas e reconhecimento de padrões são frequentemente utilizados nesse contexto.

Abordagens de Extração de Informações Estruturadas: Em alguns casos, os documentos possuem uma estrutura específica, como tabelas, formulários ou documentos XML. Nesses casos, técnicas de extração de informações estruturadas podem ser empregadas. Essas abordagens envolvem a identificação da estrutura do documento e a extração das informações relevantes de acordo com essa estrutura. Algoritmos de parsing e técnicas de expressões regulares são comumente utilizados nesse contexto.

É importante ressaltar que a extração de informações de documentos é uma área de pesquisa em constante evolução, dessa forma técnicas e algoritmos estão sendo desenvolvidos e aprimorados continuamente, impulsionados pelo avanço da inteligência artificial e do processamento de dados.

4. Desenvolvimento do framework

Um dos principais objetivos da Engenharia de Software é o reuso. Através da reutilização de software obtém-se o aumento da qualidade e redução do esforço de desenvolvimento (Gimenes & Huzita, 2005).

Com a adoção de uma solução baseada no modelo de framework, é possível contar com funcionalidades e módulos pré-desenvolvidos, como algoritmos de aprendizado de máquina, métodos de pré-processamento de dados, visualização de resultados, entre outros. Isso pode acelerar significativamente o desenvolvimento de projetos, permitindo que os desenvolvedores se concentrem mais nas especificidades do problema em questão, uma vez que o framework apresentado busca encapsular uma série de técnicas e conceitos complexos, fornecendo interfaces e ferramentas que simplificam a implementação dessas técnicas, permitindo que os desenvolvedores com diferentes níveis de experiência possam trabalhar em projetos avançados.

Além destas formas de reutilização, a tecnologia de frameworks possibilita que uma família de produtos seja gerada a partir de uma única estrutura que captura os conceitos mais gerais da família de aplicações (Pinto, 2000).

Frameworks podem ser classificados de diversas formas. Inicialmente, são classificados em dois grupos principais: frameworks de aplicação orientado a objetos (Fayad et al., 1999a, b; Fayad & Johnson, 2000) e framework de componentes. (Szyperski, 1997), que por definição, esse projeto será classificado como framework de componentes.

Os principais benefícios decorrentes da utilização de frameworks, segundo Fayad et al. (1999b) e Pinto (2000) advém da modularidade, reusabilidade, extensibilidade e inversão de controle que os frameworks proporcionam.

4.1- Processo de desenvolvimento do framework

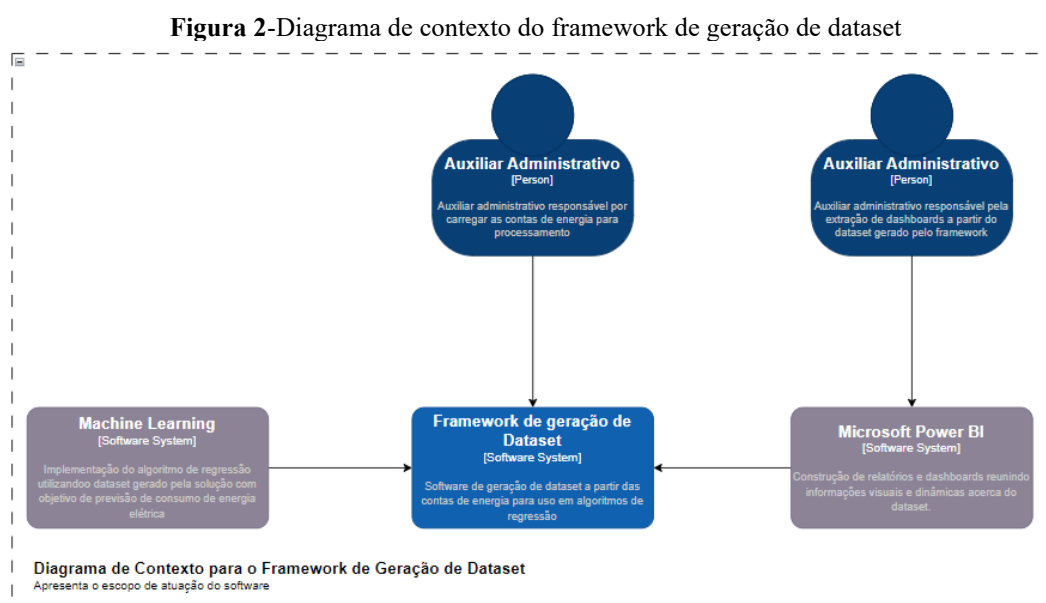
O objetivo desse framework é permitir a coleta automatizada e precisa dos dados necessários para a análise de regressão, fornecendo um dataset eficiente para prever o consumo de energia elétrica com base em variáveis identificadas na conta de energia. Neste capítulo, iremos apresentar a arquitetura de software, a arquitetura de solução em cloud, os componentes utilizados e seu fluxo básico, desde a coleta e processamento dos dados por OCR até os meios para disponibilizar seu conteúdo para uso.

Segundo Mattsson (1996, 2000), um framework é uma arquitetura desenvolvida com o objetivo de atingir a máxima reutilização, representada como um conjunto de classes abstratas e concretas, com grande potencial de especialização.

4.2- Arquitetura de software

O C4 Framework é uma abordagem eficaz para a visualização e comunicação da arquitetura de software. Ele fornece um conjunto de diagramas simples, mas poderosos, que permitem aos arquitetos de software representar as várias perspectivas de um sistema de maneira clara e concisa, começando com um contexto de alto nível e aprofundando-se em detalhes cada vez mais específicos, ajudando a organizar e estruturar informações complexas, permitindo uma compreensão rápida e eficiente do sistema.

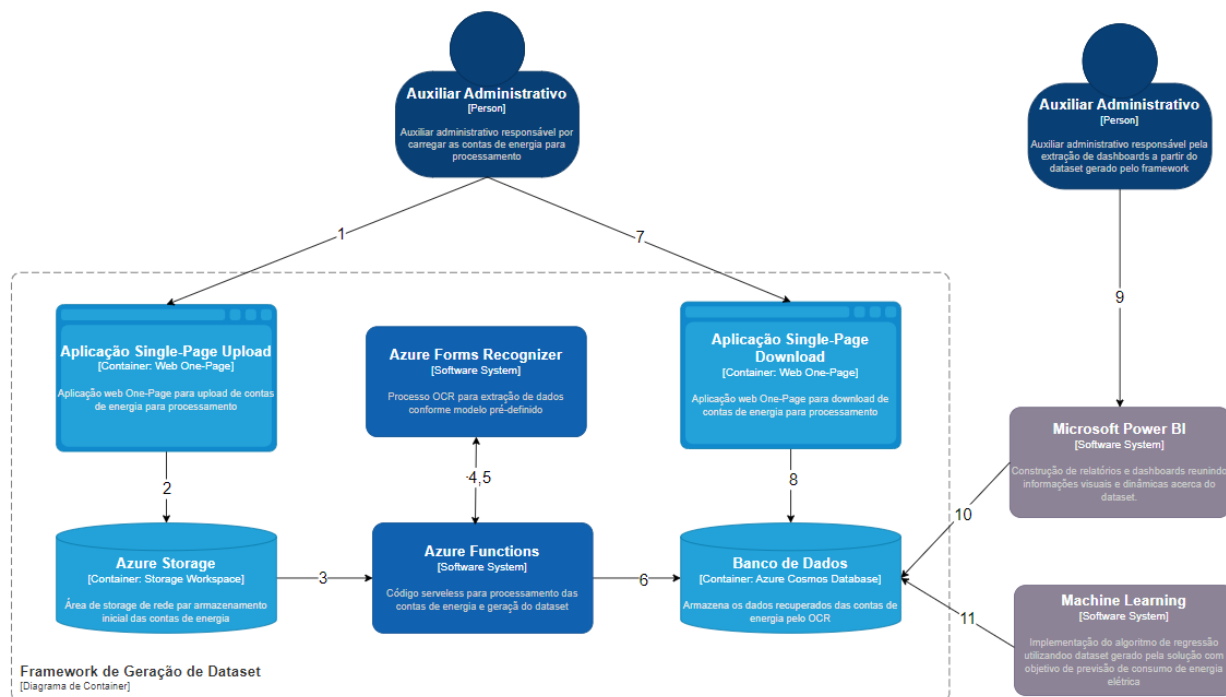
Dessa forma, na Figura 2, apresentamos a arquitetura de software através do C4 framework.



Fonte: Adaptado de Simon Brown (2006).

O diagrama do container, apresentado na Figura 3, mostra a forma de alto nível da arquitetura de software e como as responsabilidades são distribuídas por ela. Ele também mostra as principais opções de tecnologia e como os contêineres se comunicam entre si. É um diagrama simples e focado em tecnologia de alto nível que é útil para desenvolvedores de software e equipe de suporte/operações.

Figura 3-Diagrama de container do framework de geração de dataset



Fonte: Adaptado de Simon Brown (2006).

Através do diagrama de contexto, apresentado na Figura 3, temos o seguinte fluxo básico:

1-Executamos o upload das contas de energia por meio de um navegador no front-end de um aplicativo Web. Os documentos contêm imagens ou estão no formato PDF.

2-A aplicação executa armazenamento dos arquivos no Azure Storage para processamento posterior.

3-O Azure Storage dispara uma trigger para o Azure Functions que recupera o documento armazenado.

4-O Azure Functions envia o arquivo recuperado para processamento no Azure Cognitive Services, que suporta as configurações de OCR do projeto.

5-O Azure Cognitive Services executa o processamento dos arquivos iniciando a coleta automatizada dos dados por meio do OCR. Ao concluir a recuperação dos dados, Azure Cognitive Services retorna um arquivo JSON ao Azure Functions.

6-O Azure Functions após receber o arquivo JSON, inicia a segmentação desse arquivo de retorno, e em seguida executa a operação de INSERT no Banco de dados que irá armazenar a estrutura do dataset a ser produzido.

7-Via web, o usuário pode realizar o download do dataset construído pelo framework e executar o algoritmo de regressão em plataforma de sua escolha

8-A aplicação executa a recuperação de todo o dataset gerado até o momento para o usuário.

9-Via powerBI, o usuário pode visualizar dashboards e relatórios conforme sua definição a partir do dataset armazenado no banco de dados.

10-O powerBI recupera o dataset na base de dados e executa exibição conforme demanda do usuário.

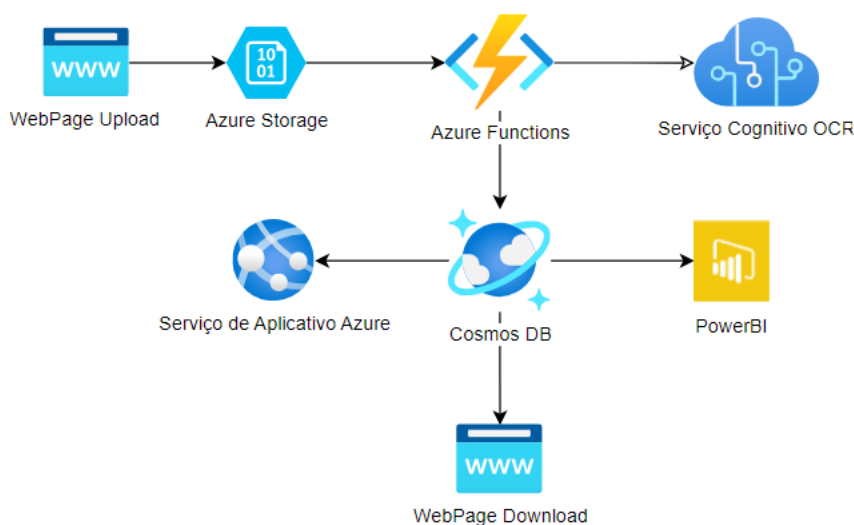
11-Outras soluções de machine learning podem se conectar a base de dados geradas pelo processo e consumir o dataset conforme suas definições.

4.3- Arquitetura de solução em nuvem

A arquitetura de computação em nuvem, é uma abordagem que envolve o projeto e a organização dos recursos de computação, armazenamento, rede e serviços em uma infraestrutura em nuvem, além de prover como vantagem a facilidade de integração e colaboração. A arquitetura de nuvem permite a interoperabilidade entre diferentes sistemas e serviços, facilitando a integração de aplicativos e o compartilhamento de dados entre eles, impulsionando a colaboração entre equipes, fornecedores e parceiros, promovendo a inovação e a agilidade nos negócios.

Na Figura 4 apresentamos o projeto e seu desenvolvimento no ambiente de nuvem Microsoft Azure.

Figura 4-Arquitetura de solução em nuvem do framework de geração de dataset



Fonte: Adaptado de The Developer's Guide to Azure (2022)

4.4- Componentes

Para a construção dessa solução no ambiente Microsoft Azure, utilizamos os seguintes componentes:

Serviço de Aplicativo é uma oferta de PaaS (plataforma como serviço) no Azure. Você pode usar Serviço de Aplicativo para hospedar aplicativos Web que podem ser dimensionados ou escalados horizontalmente manualmente ou automaticamente. O serviço dá suporte a várias linguagens e estruturas, como ASP.NET, ASP.NET Core, Java, Ruby, Node.js, PHP e Python.

Azure Functions é uma plataforma de computação sem servidor que você pode usar para criar aplicativos. Com o Functions, você pode usar gatilhos e associações para reagir a alterações nos serviços do Azure, como o Armazenamento de Blobs e o Azure Cosmos DB. As funções podem executar tarefas agendadas, processar dados em tempo real e processar filas de mensagens.

Reconhecimento de Formulários faz parte dos Serviços de IA Aplicada do Azure. Reconhecimento de Formulários oferece uma coleção de pontos de extremidade pré-criados para extrair dados de faturas, documentos, recibos, cartões de identificação e cartões de visita. Esse serviço mapeia cada parte dos dados extraídos para um campo como um par chave-valor. Reconhecimento de Formulários também extrai o conteúdo e a estrutura da tabela. O formato de saída é JSON.

O Armazenamento do Azure é uma solução de armazenamento em nuvem que inclui armazenamento de objeto, blob, arquivo, disco, fila e tabela. O Armazenamento de Blobs é um serviço que faz parte do Armazenamento do Azure. O Armazenamento de Blobs oferece armazenamento otimizado de objetos de nuvem para grandes quantidades de dados não estruturados.

O Azure Cosmos DB é um banco de dados NoSQL totalmente gerenciado, altamente responsivo e escalonável. O Azure Cosmos DB oferece segurança de nível empresarial e dá suporte a APIs para muitos bancos de dados, linguagens e plataformas. Os exemplos incluem SQL, MongoDB, Gremlin, Table e Apache Cassandra. As opções de dimensionamento automático sem servidor no Azure Cosmos DB gerenciam com eficiência as demandas de capacidade dos aplicativos.

O Power BI é uma coleção de serviços de software e aplicativos que exibem informações de análise.

5. Indicação de algoritmos de regressão

A análise de regressão é uma área fundamental na ciência de dados e estatística, permitindo a previsão e modelagem de relações entre variáveis. Existem diferentes algoritmos de regressão disponíveis, cada um com suas características e aplicabilidades específicas. A escolha correta do algoritmo pode fazer a diferença no desempenho e na precisão das previsões, dessa forma, será apresentado algumas indicações comuns de algoritmos de regressão, levando em consideração as características dos dados e os objetivos do projeto.

Conforme Hastie, T., Tibshirani, R., and Friedman, J. (2009), os modelos lineares de predição foram amplamente desenvolvidos em era pré-computador das estatísticas, mas mesmo na era atual da informática ainda são bons motivos para estudá-los e usá-los. Eles são simples e muitas vezes fornece uma descrição adequada e interpretável de como os insumos afetam a saída. Para fins de previsão, eles às vezes podem superar os mais sofisticados modelos não lineares, especialmente em situações com pequeno número de treinamento casos, baixa relação sinal-ruído ou dados esparsos.

5.1- Descrição de algoritmos de regressão

A escolha do algoritmo para estruturação do dataset depende das características dos dados e dos objetivos específicos do problema em questão. Cada algoritmo possui suas vantagens e limitações, por isso é importante selecionar aquele que melhor se adequa às particularidades dos dados e às necessidades analíticas.

Se identificado que os dados possuem uma relação linear entre as variáveis, a Regressão Linear pode ser uma escolha adequada. Por outro lado, se houver interações complexas ou não linearidades entre as variáveis, algoritmos como a Regressão de Árvore de Decisão, Regressão de Ridge ou até mesmo modelos de aprendizado de máquina mais avançados, como Redes Neurais Artificiais, podem ser mais indicados. Nesse cenário deve-se também considerar o tamanho do dataset, a presença de ruídos ou outliers, a interpretação do modelo e outros fatores relevantes para a análise. Um cuidadoso estudo exploratório dos dados pode ajudar na identificação de padrões e na escolha do algoritmo mais apropriado.

Dessa forma, a escolha do algoritmo para estruturação do dataset é uma etapa crucial no processo de análise de regressão. É fundamental considerar as características dos dados e os

objetivos específicos do problema, garantindo a seleção do algoritmo mais adequado para obter resultados precisos e significativos. Na Tabela 2 apresentamos uma breve descrição sobre os tipos de algoritmos que sugerimos para uso do dataset gerado pelo framework.

Tabela 2-Algoritmos de regressão para uso do dataset

ALGORITMO	DESCRIÇÃO
Algoritmo de Regressão Linear	A regressão linear é um algoritmo simples e amplamente utilizado para problemas de regressão. Ele assume uma relação linear entre as variáveis dependentes e independentes. Se os dados de consumo de energia elétrica exibirem um padrão de crescimento linear ao longo do tempo, a regressão linear pode ser uma escolha adequada.
Árvores de Decisão	As árvores de decisão são algoritmos de aprendizado de máquina que dividem o conjunto de dados em subconjuntos menores com base em regras de decisão. Elas são capazes de lidar com dados não lineares e podem capturar relações complexas entre as variáveis. As árvores de decisão são particularmente úteis se o consumo de energia elétrica for influenciado por uma combinação de fatores sazonais, demográficos e meteorológicos.
Random Forest	Random Forest é um algoritmo baseado em conjunto que utiliza várias árvores de decisão para realizar a regressão. Ele combina as previsões de várias árvores para obter uma previsão mais robusta e precisa. Random Forest é uma escolha recomendada quando há um grande número de variáveis de entrada ou quando há interações complexas entre elas.
Redes Neurais	Redes neurais são algoritmos poderosos para lidar com problemas de regressão. Elas são capazes de aprender relações não lineares complexas e podem se adaptar a diferentes tipos de dados. Redes neurais podem ser uma boa escolha se houver uma quantidade considerável de dados disponíveis para treinamento e se o problema de consumo de energia elétrica for altamente não linear.

Fonte: Adaptado de Éliton Fontana (2020).

É importante realizar experimentos e avaliar o desempenho de diferentes algoritmos em relação às métricas de avaliação relevantes, como erro médio quadrático (RMSE) ou coeficiente de determinação (R^2). Além disso, considere a interpretação do modelo, a complexidade computacional e a escalabilidade em relação ao tamanho do dataset.

A escolha do algoritmo para estruturação do dataset depende das características dos dados, dos objetivos do problema e das métricas de avaliação relevantes. É recomendado realizar testes comparativos para selecionar o algoritmo que melhor se ajuste ao contexto do consumo mensal de energia elétrica.

6. Resultados e aplicações

6.1- Principais aplicações práticas

A previsão de consumo elétrico tem várias aplicações práticas importantes em diferentes setores. Na Tabela 3 destacamos os temas em que a previsão de consumo pode se destacar.

Tabela 3 - Principais aplicações práticas da previsão de consumo de energia elétrica

TEMA	ESCOPO
Gestão de Energia Residencial	A previsão de consumo elétrico pode ser usada em residências para auxiliar na gestão de energia. Com base nas previsões, os consumidores podem ajustar seu comportamento de consumo, programar e otimizar o uso de aparelhos elétricos e tomar decisões informadas para economizar energia e reduzir os custos
Planejamento da Demanda	As empresas de distribuição de energia precisam planejar a infraestrutura e os recursos necessários para atender à demanda de energia em uma determinada região. A previsão de consumo elétrico permite estimar a demanda futura com base em dados históricos e fatores sazonais, facilitando o planejamento adequado do fornecimento de energia
Operação de Redes Elétricas	As empresas de distribuição de energia também podem usar a previsão de consumo elétrico para otimizar a operação das redes elétricas. Com base nas previsões, elas podem equilibrar a geração e distribuição de energia, evitar sobrecargas e implementar estratégias de gerenciamento de carga para garantir um fornecimento eficiente e confiável.
Comercialização de Energia	No setor de comercialização de energia, a previsão de consumo elétrico desempenha um papel fundamental. Com base nas previsões, os agentes de mercado podem tomar decisões de compra e venda de energia, otimizar a alocação de recursos e minimizar os riscos associados à flutuação do mercado.
Planejamento de Energia Renovável	A previsão de consumo elétrico é crucial para o planejamento e a integração de fontes de energia renovável, como energia solar e eólica. As previsões ajudam a determinar a capacidade necessária de geração renovável, otimizar a combinação de fontes de energia e planejar a infraestrutura de armazenamento de energia para lidar com as variações de geração dessas fontes.
Eficiência Energética em Indústrias	As indústrias podem usar a previsão de consumo elétrico para identificar padrões e tendências em seu consumo de energia. Com base nessas previsões, elas podem implementar medidas de eficiência energética, como ajuste de horários de produção, controle de carga e adoção de tecnologias mais eficientes, visando reduzir custos operacionais e melhorar a sustentabilidade.

Fonte: Adaptado de Junior (2007)

Essas são apenas algumas das principais aplicações práticas da previsão de consumo elétrico. É uma área em constante evolução, impulsionada pela crescente demanda por eficiência energética, integração de energias renováveis e tomada de decisões informadas no setor de energia.

6.2-Análise sobre a qualidade e confiabilidade das informações extraídas das faturas de energia elétrica

A análise das faturas de energia elétrica é uma prática comum entre consumidores e empresas que buscam entender seus gastos e tomar medidas para otimizar o consumo de energia. No entanto, a qualidade e confiabilidade das informações presentes nessas faturas têm sido objeto de discussão. Neste texto, examinaremos os principais pontos relacionados a esse tema, explorando os desafios associados à precisão das informações contidas nas faturas de energia elétrica.

A qualidade das informações presentes nas faturas de energia elétrica é crucial para que os consumidores possam tomar decisões informadas sobre seu consumo e compreender as origens de seus custos energéticos. No entanto, existem alguns desafios que podem afetar essa qualidade. Um deles é a precisão dos medidores utilizados para registrar o consumo de energia. Falhas ou desvios nesses dispositivos podem resultar em informações incorretas sendo refletidas nas faturas. Outro fator que pode influenciar a qualidade das informações é a complexidade da própria estrutura das faturas de energia elétrica. Muitas vezes, elas contêm uma série de componentes, como tarifas, impostos, encargos setoriais e bandeiras tarifárias, o que pode dificultar a compreensão das informações pelo consumidor.

A confiabilidade das informações nas faturas de energia elétrica é outro aspecto importante que suscita discussões. Os consumidores podem questionar a precisão dos dados fornecidos pelas empresas de energia elétrica, especialmente quando há variações significativas no valor cobrado de um mês para o outro. Além disso, surgem questionamentos sobre a possibilidade de erros humanos durante o processo de leitura e emissão das faturas, o que pode resultar em informações inconsistentes ou incorretas.

Dessa forma, medidas para garantir a qualidade e confiabilidade das informações têm sido implementadas para garantir a qualidade e confiabilidade das informações presentes nas faturas

de energia elétrica. Uma delas é a regulamentação governamental que estabelece padrões para aferição e verificação dos medidores de energia. Essas normas visam assegurar que os dispositivos utilizados sejam precisos e estejam em conformidade com os requisitos técnicos.

A discussão sobre a qualidade e confiabilidade das informações extraídas das faturas de energia elétrica é relevante, pois afeta a capacidade dos consumidores de entender e controlar seus gastos energéticos. Embora existam desafios associados à precisão e confiabilidade dessas informações, medidas estão sendo adotadas para garantir a conformidade com padrões regulatórios e aprimorar os sistemas de medição e faturamento. É fundamental que os consumidores acompanhem suas faturas, questionem eventuais discrepâncias e busquem informações adicionais junto às empresas de energia elétrica para tomar decisões mais informadas e eficientes em relação ao consumo de energia

7-Conclusão

Nos últimos anos, tem havido um aumento significativo no consumo de energia elétrica em todo o mundo. Fatores como crescimento populacional, desenvolvimento econômico, urbanização e proliferação de dispositivos eletrônicos têm desempenhado um papel importante no aumento do consumo de energia elétrica, no entanto, é importante ressaltar que o aumento do consumo de energia elétrica também traz preocupações ambientais, devido ao uso de fontes de energia não renováveis, como carvão e gás natural, que contribuem para a emissão de gases de efeito estufa e mudanças climáticas.

Nesse contexto, temas como eficiência energética, energia renovável, armazenamento de energia e gestão inteligente de energia se tornaram relevantes no contexto atual e futuro, e, apesar de temas distintos, todos visam o mesmo objetivo: o uso de tecnologias e práticas que reduzem a quantidade de energia necessária para realizar uma determinada tarefa.

Considerando o tema de gestão inteligente de energia, o presente trabalho, através do uso de OCR e técnicas de processamento de imagem permite automatizar a extração de informações relevantes das faturas de energia elétrica, eliminando a necessidade de entrada manual de dados, economizando tempo e esforço significativos, proporcionando uma maneira precisa e consistente de extrair informações-chave das faturas, evitando erros humanos e melhorando a qualidade dos dados utilizados para o cálculo do consumo mensal de energia elétrica, podendo ser aplicado a um grande volume de faturas de energia elétrica, permitindo o processamento eficiente de dados em lote ou em tempo real.

Isso é especialmente relevante em cenários onde há um grande número de clientes ou faturas para processar. Com o uso de um algoritmo de regressão treinado com base no dataset estruturado, gerado por essa solução, pode levar a previsões mais precisas e confiáveis do consumo mensal de energia elétrica, auxiliando as empresas a planejar melhor o custo de energia, otimizando a capacidade e melhorando a eficiência operacional, além da previsão de consumo.

No entanto, é importante considerar que a construção desse framework requer conhecimentos especializados em processamento de imagem, OCR, aprendizado de máquina e algoritmos de regressão. Além disso, a privacidade e a segurança dos dados dos clientes devem ser tratadas com cuidado, garantindo a conformidade com as regulamentações de proteção de dados.

Referências

Weizenbaum, Joseph. **ELIZA - A Computer Program for the Study of Natural Language Communication between Man and Machine**. Association for Computing Machinery, p. 36-45. Disponível em: <https://dl.acm.org/doi/pdf/10.1145/365153.365168> . Acesso em: 24 de fevereiro de 2024.

Brown, Simon (2006). **C4 Model for Visualizing Software Architecture**. Leanpub. Disponível em: <https://s3.amazonaws.com/static.codingthearchitecture.com/presentations/aotb2019-visualising-software-architecture-with-the-c4-model.pdf> . Acesso em: 24 de fevereiro de 2024.

Altaiar, H. Garverick, J., Babe, I., Toroman, M., Lee J., Minasyan, V (2022).. The Developer's Guide to Azure. Microsoft E-Book series. Disponível em: <https://clouddamcdnprodep.azureedge.net/gdc/gdcz5VglF/original?ocid=AID3042111> . Acesso em: 24 de fevereiro de 2024.

Hastie, T., Tibshirani, R., and Friedman, J. **The Elements of Statistical Learning: Data Mining, Inference, and Prediction**. Springer series in statistics, Springer, 2009.

Gimenes, I. M. S., Huzita, E.H.M. **Desenvolvimento Baseado em Componentes: Conceitos e Técnicas**. Ciência Moderna, Rio de Janeiro, 2005.

Pinto, S. C. C. S. **Composição em Web Frameworks**. Tese (Doutorado). Departamento de Informática PUC-Rio, Rio de Janeiro, 2000.

Fayad, M.E., Johnson, R.E. **Domain-specific application frameworks: frameworks experience by industry**. John. Wiley, New York, 2000.

Fayad, M.E., Schimdt, D.C., Johson, R.E. **Building application frameworks: object-oriented foundations of framework desing**. John Wiley, New York, 1999a.

Fayad, M.E., Schimdt, D.C., Johson, R.E. **Implementing Application Frameworks: Object-Oriented Frameworks at Work**. John Wiley, New York, 1999b.



XII SIMPÓSIO DE ENGENHARIA DE PRODUÇÃO

“A contribuição da Engenharia de Produção para alcance dos Objetivos de Desenvolvimento Sustentável da ONU.”

Rio de Janeiro, Rio de Janeiro, Brasil – 24 a 26 de Maio de 2024.

Szyperski, Clemens. **Component Software. Beyond Object-Oriented Programming**. Pearson, Londres, 2011.

Mattsson, M. **Evolution and Composition Object-Oriented Frameworks**. Tese (Pós-Doutorado). University of Karskrona/Ronneby, Department of Software Engineering and Computer Science, Suécia, 2000.

Mattsson, M. **Object-Oriented Frameworks – A Survey of Methodological Issues**. Tese (Mestrado). Department of Computer Science, Lund University, Suécia, 1996.

Junior H.Q.P. **Economia da Energia – Fundamentos Econômicos, Evolução Histórica e Organização Industrial**. LTC, rio de Janeiro, 2007.