



DANILO MOTA FERREIRA LEMES

RELATÓRIO FINAL DE ESTÁGIO SUPERVISIONADO

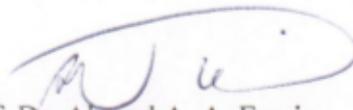
LAVRAS - MG

2019

DANILO MOTA FERREIRA LEMES

RELATÓRIO FINAL DE ESTÁGIO SUPERVISIONADO

Relatório de estágio supervisionado
apresentado a Universidade Federal
de Lavras, como parte das exigências
do curso de Sistemas de Informação,
para a obtenção do título de Bacharel.



Prof. Dr. Ahmed A. A. Esmín
Orientador

LAVRAS - MG

2019

DANILO MOTA FERREIRA LEMES

RELATÓRIO FINAL DE ESTÁGIO SUPERVISIONADO

RESUMO

O presente relatório descreve as atividades desenvolvidas durante o período em que atuei como estagiário para a empresa LCS Link Engenharia, sediada na cidade de São Paulo - SP.

A LCS Link é uma empresa especializada em Engenharia Elétrica, atuando nos mercados de Missão crítica, Geração e Transmissão de Energia, oferecendo uma grande gama de soluções e serviços para os clientes desses segmentos. Minha atuação como estagiário na empresa aconteceu como membro de uma equipe interna, situada em Lavras - MG, especializada em desenvolvimento de software e inteligência artificial, trabalhando em cima das demandas fechadas pela filial em São Paulo.

O estágio teve duração de cerca de 6 meses, e visava a execução de estudos e projetos voltados ao desenvolvimento de modelos de aprendizagem de máquina e o processamento de dados internos da empresa.

Considero esta uma experiência enriquecedora, permitindo contato com a realidade de um ambiente de desenvolvimento real, além de aprender e entender melhor as tecnologias e conceitos aplicados no dia a dia dos profissionais da área.

ABSTRACT

This report describes the activities developed during the period in which I worked as a trainee for the company LCS Link Engenharia, headquartered in the city of São Paulo - SP.

LCS Link is a company that specializes in Electrical Engineering, acting in the markets of Critical Mission, Generation and Transmission of Energy, offering a wide range of solutions and services to customers in these segments. My internship at the company happened as a member of an internal team, located in Lavras - MG, specialized in software development and artificial intelligence, working on the demands closed by the branch in São Paulo.

The internship lasted for about 6 months, and focused on the execution of studies and projects aimed at the development of machine learning models and the internal data processing of the company.

I consider this an enriching experience, allowing contact with a real development environment, in addition to learning and better understanding the technologies and concepts applied in the day to day of the professionals in the area.

SUMÁRIO

1	INTRODUÇÃO.....	1
2	REFERENCIAL TEÓRICO.....	2
2.1	IoT.....	2
2.2	Aprendizagem de máquina.....	2
2.3	Clusterização.....	3
2.4	Detecção de outliers.....	4
3	DESCRIÇÃO DA ORGANIZAÇÃO.....	5
3.1	LCS Link.....	5
3.2	Tecnologias.....	6
3.2.1	Trello.....	6
3.2.2	Python.....	6
3.2.3	Jupyter Notebook.....	7
3.2.4	Node-Red.....	7
4	ATIVIDADES DESENVOLVIDAS.....	9
4.1	Modelo de monitoramento de sensores de energia.....	9
4.2	Configuração de ambiente colaborativo de desenvolvimento.....	11
4.3	Módulo de processamento de requisições.....	12
5	CONCLUSÃO.....	15

1 INTRODUÇÃO

Ao me aproximar do final do curso de sistemas de informação, fui convidado pelo meu orientador a realizar uma entrevista para estagiar como analista de dados em uma equipe trabalhando para a LCS Link Engenharia, vi isto como uma oportunidade única, pois gostaria de alguma experiência real antes de terminar a graduação e ser jogado no mercado, realizei então a entrevista na esperança de ser chamado e assim fui.

Este relatório procura descrever as atividades desenvolvidas e introduzir a empresa e equipe na qual atuei durante o estágio, apresentando as principais tecnologias com as quais tive contato ao longo desse período e explicando alguns conceitos primários da área.

Como estagiário na área de desenvolvimento de software e machine learning em uma equipe pequena, tive contato direto com meu orientador durante os projetos, atuando de forma a dar suporte para as atividades da equipe. As atividades a mim designadas durante o período de estágio contemplavam a análise, coleta e visualização de dados, desenvolvimento de aplicações para processamento de informações no backend assim como estudo e desenvolvimento de modelos de aprendizagem de máquina para implementação futura no ambiente de monitoramento desenvolvido internamente.

Ao final do relatório, deve-se ter uma visão geral da empresa e tecnologias com as quais tive contato, assim como das atividades realizadas durante o período em que atuei como estagiário.

2 REFERENCIAL TEÓRICO

Para que exista uma melhor compreensão da área de atuação da empresa e das atividades realizadas durante o estágio, é necessário primeiramente uma base de conhecimento inicial em relação aos conceitos e terminologias empregadas.

Ao fim deste capítulo, teremos os principais conceitos necessários para o entendimento do trabalho.

2.1 IoT

Atualmente, não só computadores e smartphones possuem capacidade de conexão a internet, inúmeros outros objetos do dia a dia vêm ganhando esta capacidade ao longo dos anos, TV's, Smartwatches, eletrodomésticos e outros vêm aderindo a tendência de um mundo cada vez mais conectado.

Através da capacidade de comunicação dentro de uma rede, equipamentos podem compartilhar informações entre si e com serviços construídos em cima dessa infraestrutura, surgindo toda uma nova área para desenvolvimento de aplicações, os chamados ambiente inteligentes (MANCINI, 2018).

A internet das coisas pode ser definida, então, como a conexão dos objetos a internet. É uma rede de objetos físicos conectados (com capacidade computacional), capazes de receber e transmitir dados, permitindo sua utilização como provedores de serviços e transmissores de informação. Com a vasta gama de novos aparelhos conectados, surgem inúmeras possibilidades de aplicações que possam se aproveitar dos dados gerados por esses equipamentos, com aplicações acadêmicas e para o mercado.

2.2 Aprendizagem de máquina

A aprendizagem de máquina é um subcampo da ciência da computação que surgiu a partir do reconhecimento de padrões na área de Inteligência Artificial, englobando o processo de estudo, análise e criação de algoritmos que consigam aprender a partir de dados categorizados previamente e, a partir destes, realizar previsões em novos conjuntos de dados (SIMON et al, 2016).

Os algoritmos de aprendizagem de máquina são classificados primariamente em duas categorias:

- Aprendizagem supervisionada
O algoritmo é treinado a partir de uma base de dados categorizada que serve como exemplo para as predições a serem realizadas.
- Aprendizagem não-supervisionada
O algoritmo recebe um conjunto de dados e deve encontrar as relações e padrões automaticamente.

A aplicação dos conceitos de aprendizagem de máquina é útil em projetos com grande volume de dados a serem analisados, padrões de difícil observação e que requerem automatização de tarefas baseadas em dados, encontrando informações que talvez passassem despercebidas aos olhos humanos e agilizando processos e respostas.

2.3 Clusterização

Conjunto de técnicas que procura realizar agrupamento automático de dados baseados em suas características. É baseada na análise das diferenças entre objetos e utiliza funções de regulação de distância para realizar a classificação do modelo (YOUGUO LI, 2016).

Diferentemente do conceito de classificação, a Clusterização não realiza nenhuma suposição a respeito dos grupos, realizando uma forma de treinamento não-supervisionado a partir dos dados analisados e classificando-os de acordo com o número de clusters definidos para a execução (valor k).

Através da análise das características, elementos são classificados como pontos em um plano, de acordo com seu grau de semelhança. Após a classificação, a função de distância entre os agrupamentos de pontos pode servir como a medida de similaridade dos padrões analisados, com elemento semelhantes sendo localizados próximos um ao outro.

2.4 Detecção de outliers

A detecção de outliers , também conhecida como detecção de novidades, detecção de anomalias, detecção de ruído ou mineração de exceção, abrange um conjunto de técnicas que procura identificar elementos incomuns dentro de um grupo de amostras analisadas.

Como definido por Barnett e Lewis (1994): Um outlier é uma observação (ou subconjunto de observações) que parece ser inconsistente com o restante daquele conjunto de dados.

É imprescindível, em tarefas de monitoramento, detectar uma mudança brusca no padrão dos dados observados, evitando a contaminação de datasets, indicando falhas de processo, erros de hardware e leituras incorretas. A detecção de outliers realiza isso analisando e comparando as séries temporais de estatísticas de uso (J. HODGE, 2004).

3 DESCRIÇÃO DA ORGANIZAÇÃO

As informações sobre a organização onde atuei como estagiário, área de atuação e as principais tecnologias envolvidas no dia a dia são apresentadas a seguir.

3.1 LCS Link

A LCS Link é uma empresa especializada no fornecimento de serviços na área de automação industrial de energia elétrica, com sede na cidade de São Paulo, e atuante no mercado a 21 anos. Possui entre seus clientes bancos, data centers, usinas hidrelétricas e outros, oferecendo soluções para as demandas dos seguintes segmentos:

- Geração e Transmissão de energia

Oferece serviços como Automação para grupos geradores e usinas hidroelétricas, Integração e desenvolvimento de software na área de PLC (Power Line Communications) e soluções para comunicação de dados, oferecendo suporte para demandas de clientes nas indústrias de energia em geral.

- Missão crítica

Missão crítica se refere à disponibilidade de aplicações e serviços para processos em que a paralisação implica em perdas financeiras e sociais. A LCS oferece diversas soluções na área, como Desenvolvimento de BMS's (Building Management Systems), Instalação de Centros de controles, criação de Sistemas de Gerenciamento dos Operadores (SGO), entre outros.

Apesar do mercado específico de atuação, a LCS Link conta com uma vasta gama de clientes, e sua atuação a anos na área de missão crítica reflete a qualidade e comprometimento da empresa com seus clientes.

A equipe em que atuei como estagiário está locada em Lavras, e foi criada para atender internamente as demandas específicas da área de desenvolvimento de software, inteligência computacional e internet das coisas (IoT), dentro dos projetos fechados pela

empresa. Esta equipe trabalha constantemente para a implementação de novas técnicas e funcionalidades para agregar valor aos serviços prestados.

3.2 Tecnologias

Durante meu período como estagiário, tive contato com vários conceitos e tecnologias da área, aquelas que mais se destacaram e, com as quais tive maior contato durante minha atuação são descritas a seguir:

3.2.1 Trello

O Trello é uma ferramenta de gerenciamento de tarefas e projetos disponível em versão web e para dispositivos móveis, desenvolvido para ter fácil usabilidade, sendo adotado pela equipe para gerenciamento das atividades do dia a dia.

A ferramenta utiliza três elementos básicos para criar uma interface simples e de fácil entendimento, sendo estes elementos, o quadro, responsável por conter as tarefas de um projeto, listas categorizadas de atividades e cards, que armazenam as informações de uma atividade específica.

3.2.2 Python

Python é uma linguagem de programação interpretada, orientada a objetos e utilizada para programação geral. Foi criada inicialmente por Guido Van Rossum em 20 de fevereiro de 1991 e disponibilizada como um projeto de código aberto, contando com milhares de colaboradores ao redor do mundo nos dias de hoje.

Possuindo uma semântica de fácil assimilação, é utilizada nos mais diversos projetos e aplicações, servindo como base para serviços web, computação científica, desenvolvimento de games entre outros. Conta também com uma extensa biblioteca de módulos disponibilizados pela comunidade de desenvolvedores para aplicação nas mais diversas áreas do conhecimento.

Devido ao fato de não possuir uma etapa de compilação, possui um ciclo de escrita-teste-depuração rápido, oferecendo maneiras simples de localizar erros no código

e contando com suporte a breakpoints e um depurador poderoso para casos mais complexos.

3.2.3 Jupyter

Jupyter é um ambiente computacional interativo para criação e edição de notebooks Jupyter, um tipo de documento estruturado em células contendo código, tabelas, mídia e texto, permitindo execução sequencial, possibilitando a fácil reprodução de código e visualização de análises. É reconhecidamente uma das tecnologias padrões na computação científica.

É um projeto de código aberto, com alta extensibilidade de funções através de plugins e suporte a execução de várias linguagens de programação através da instalação de Kernels para interpretação de código. Seu nome é uma referência as três linguagens suportadas inicialmente, Julia, Python e R.

Atualmente o projeto é responsável pelo desenvolvimento e manutenção de três produtos de computação interativa, Jupyter Notebook, o projeto inicial, JupyterHub, um servidor multi-usuários para armazenamento de notebooks na nuvem, e JupyterLab, uma versão atualizada do JupyterNotebook, com mais funcionalidades e interface otimizada.

3.2.4 Node-Red

Ferramenta de desenvolvimento baseada em fluxo criada pela IBM, construído em cima do ambiente Node.js para interconectar aparelhos de hardware, serviços e API's, voltada a rede de internet das coisas (IoT), permitindo a automação de requisições, execuções de scripts e chamadas de funções de forma relativamente simples.

A tecnologia oferece programação visual de funções posteriormente convertidas em JavaScript, e vem pré instalada de fábrica em vários modelos de remotas como o Raspberry Pi, Intel IoT Gateway, entre outros.

A programação na ferramenta acontece através de blocos, que representam funções predeterminadas, podendo ser configuradas e interligadas de acordo com a necessidade do programador. Existem blocos para execução de código próprio, Triggers para execução em intervalos predeterminados, Requisições POST e GET entre muitos outros

suportados por padrão. Existindo ainda a possibilidade de instalação de blocos adicionais criado pela comunidade, estendendo as funcionalidades da tecnologia.

4 ATIVIDADES DESENVOLVIDAS

Durante o período em que atuei como estagiário, tive a oportunidade de trabalhar e desenvolver diferentes atividades dentro do ambiente de produção, atuando em projetos de forma a prestar suporte as funcionalidades implementadas pela equipe de desenvolvedores.

4.1 Modelo de monitoramento de sensores de energia

Como parte da equipe de dados, trabalhei em um projeto para monitoramento e estudo de sensores de energia, procurando analisar padrões de consumo de eletricidade e suas variações, para a criação de um modelo preditivo posteriormente, sendo este capaz de definir se um aparelho está ligado ou desligado, com base nas leituras de energia e também detectar outliers no padrão de consumo observado.

Para monitoramento e teste, foram utilizados os sensores disponíveis no próprio laboratório de desenvolvimento da equipe, com foco no comportamento das leituras energéticas do sensor instalado no ar condicionado do ambiente, devido ao alto consumo de aparelhos desse tipo. Os sensores são conectados diretamente a fiação elétrica do equipamento a ser analisado e também a um microcontrolador (remota) que realiza a distribuição de diferentes fontes e sensores, os dados retornados são geralmente strings representativas da grandeza medida, com sensores de energia retornando a corrente medida para o aparelho.

A etapa inicial consistiu no levantamento dos tipos de dados e a melhor forma de tratá-los, onde a captura referente ao gasto energético deveria cobrir os intervalos de uso durante dias úteis, e onde ambos estados (Ligado & Desligado) para o aparelho em questão deveriam estar representados, durante diferentes períodos do dia em diferentes dias.

A aquisição destas informações foi realizada através do banco de dados próprio, onde existem tabelas referentes ao consumo histórico dos sensores analisados. Para casos onde não exista essa possibilidade de acesso ao histórico via banco de dados, geralmente em fases iniciais de prototipação, implementei também uma solução para captura de dados de energia via requisições ao IP da remota com os dados do sensor.

A segunda etapa do projeto lidou com a análise dos dados obtidos e o teste e preparação do modelo.

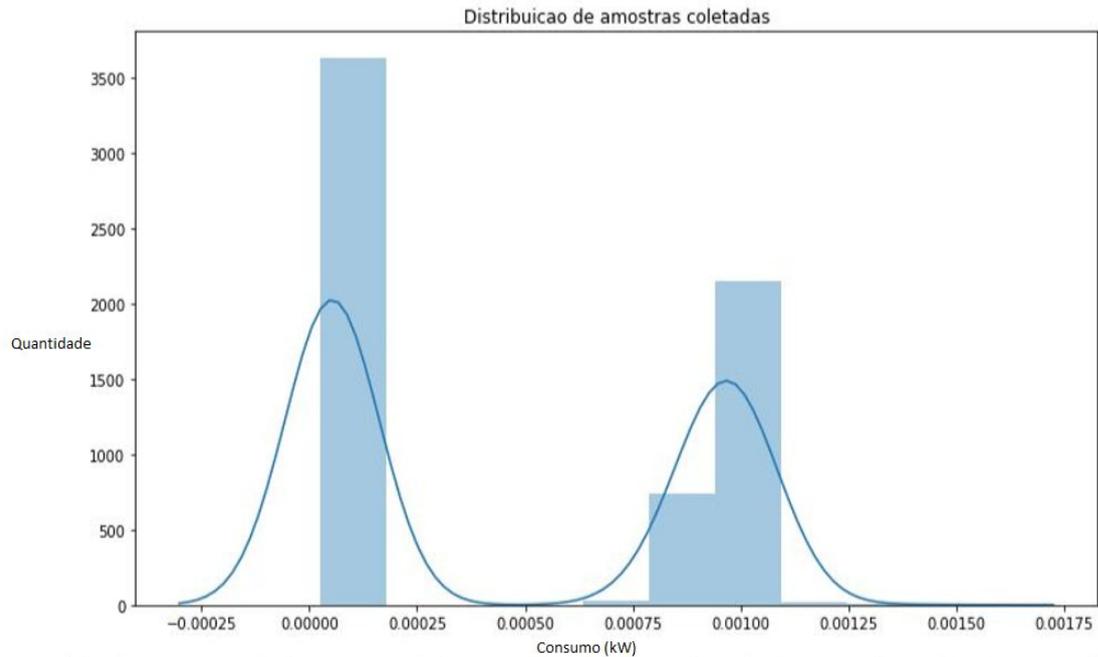


Figura 1. Visualização de amostras coletadas

Devido às características energéticas apresentadas pelos dados monitorados, pode se notar na figura 1 uma clara diferença no consumo entre os estados possíveis, optando-se assim pela utilização de um algoritmo de clusterização KMeans como forma de identificação dos grupos energéticos e seus limites, dada a diferença dos valores energéticos entre os mesmos (On & Off). A vantagem da utilização do método se deve ao fato do mesmo não necessitar de labels nos dados de treinamento, e tomando como entrada os valores energéticos coletados, é capaz de analisar e deduzir os clusters sem ajuda externa, ideal para identificação dos padrões de consumo de diferentes aparelhos e definir seus estados energéticos.

Através dos clusters foi possível então identificar os maiores e menores valores de consumo, e então delimitar uma região intermediária de threshold entre os estados, definida através do cálculo das distâncias entre o maior ponto registrado no Cluster 0 (Desligado) e Cluster 1 (Ligado) para definir quando o aparelho está desligado ou ligado.

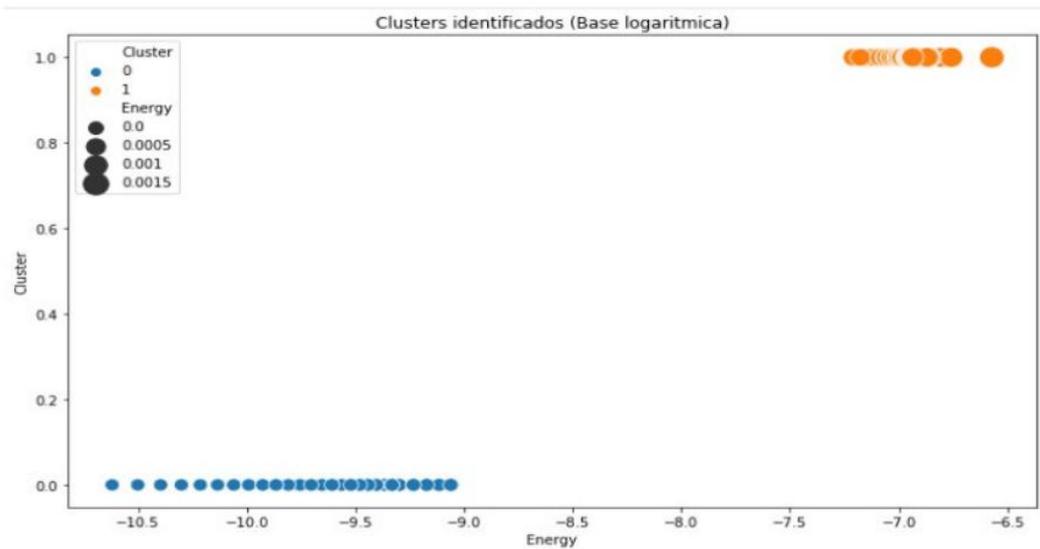


Figura 2. Visualização dos clusters de consumo para cada estado.

Com os clusters definidos (Figura 2), foi possível então trabalhar em um modelo para detecção de outliers no consumo registrado, picos de consumo que ocorrem naturalmente durante a leitura de energia, que não são representantes do consumo real. Depois de uma análise das alternativas, foi escolhido o algoritmo de identificação de outliers Isolation Forest para ser aplicado no modelo.

Após a utilização da clusterização, o modelo Isolation Forest foi então treinado separadamente para os estados (ON & OFF) utilizando-se de uma coluna adicional no dataset, adicionada pelo algoritmo de cálculo de threshold criado. O treinamento acontece separadamente devido a alta variação energética entre os dois estados, permitindo assim classificações mais precisas em relação a outliers.

4.2 Configuração de ambiente colaborativo de desenvolvimento

Atuando na equipe de dados, constantemente existia uma necessidade de compartilhamento de documentos, relatórios e códigos com meu coordenador. Devido a este motivo, surgiu a necessidade de encontrar um software/serviço intermediário para facilitar e agilizar esse processo. Fui então incumbido da tarefa de pesquisa a solução que melhor se adaptasse ao nosso ambiente de desenvolvimento.

As informações de um projeto podem ser compartilhadas de inúmeras maneiras, porém, em um ambiente de produção é necessário certa padronização, além do controle

de documentação, dados gerados e a garantia das mesmas condições de execução de um script, entre a máquina do desenvolvedor e do responsável pelo projeto.

Visando atender as necessidades de compartilhamento e o potencial de crescimento da equipe no futuro, as principais características a serem consideradas na pesquisa pela ferramenta eram a capacidade de atender múltiplos usuários, ser multi-projetos, e a disponibilidade da aplicação em código aberto.

Após pesquisa, as seguintes ferramentas foram consideradas:

1. JupyterHub
2. Apache Zeppelin
3. SageMathInc COCALC
4. BinderHub

A avaliação foi realizada com instalações em ambientes virtuais das quatro ferramentas, destacando-se neste processo o ambiente COCALC, contando com funcionalidades adicionais em comparação aos outros candidatos, como suporte a chat entre usuários da aplicação, edição simultânea de arquivos e rollback para versões anteriores de arquivos.

Inicialmente, um projeto voltado a aplicações matemáticas, O COCALC é uma aplicação Web que sustenta um ambiente virtual colaborativo para criação e edição de códigos e documentos de uma vasta gama de formatos, possuindo suporte nativo as linguagens de formatação como LaTeX e Markdown, as principais ferramentas para estatística e Data Science (Pandas, Scikit-Learn, TensorFlow, Jupyter), e linguagens como Python, R, Julia, entre outros, suportando ainda a adição de novos interpretadores através de plugins.

Depois de definido o ambiente a ser utilizado, realizei a instalação da aplicação em uma máquina definida como servidor para alguns serviços de teste da equipe, o processo foi realizado sem problemas através de uma imagem Docker, e seguindo as instruções fornecidas pelos desenvolvedores.

4.3 Módulo de processamento de requisições

No projeto mais recente em que trabalhei no meu período como estagiário, desenvolvi um módulo Python, para utilização interna, com o objetivo de processar a resposta dos

endpoints internos mantidos pela equipe, responsáveis pela requisição de dados para o sistema de monitoramento desenvolvido internamente. O ambiente de monitoramento se resume em uma aplicação web em formato de dashboard que permite a assimilação e visualização dos dados de ambientes monitorados pela equipe.

A motivação para este projeto surgiu da necessidade de obtenção de informações processadas e formatadas para apresentação em um protótipo de chatbot em desenvolvimento na empresa. As chamadas aos endpoints retornam muitas vezes conjunto de dados, listas e dicionários de difícil entendimento, e servindo como um assistente automatizado para os clientes dos sistemas de monitoramento, é ideal que o cliente chatbot tenha o retorno de suas requisições no melhor formato de texto desejado para visualização em chat.

O projeto consiste em um módulo principal, responsável pelo login e gerenciamento do token de sessão, renovando-o quando necessário, e módulos específicos para cada categoria de informação a ser processada, as categorias se resumem em informações de consumo de aparelhos assim como previsões para os mesmos, informações do ambiente monitorado, como alertas e erros que podem indicar falhas no sistema e informações gerais do ambiente como temperatura e umidade do local.

O processamento consistia em recuperar objetos JSON retornados pela API, realizar a verificação das chaves do objeto, e formatar Strings a serem retornadas para o cliente chatbot de acordo com seu conteúdo.

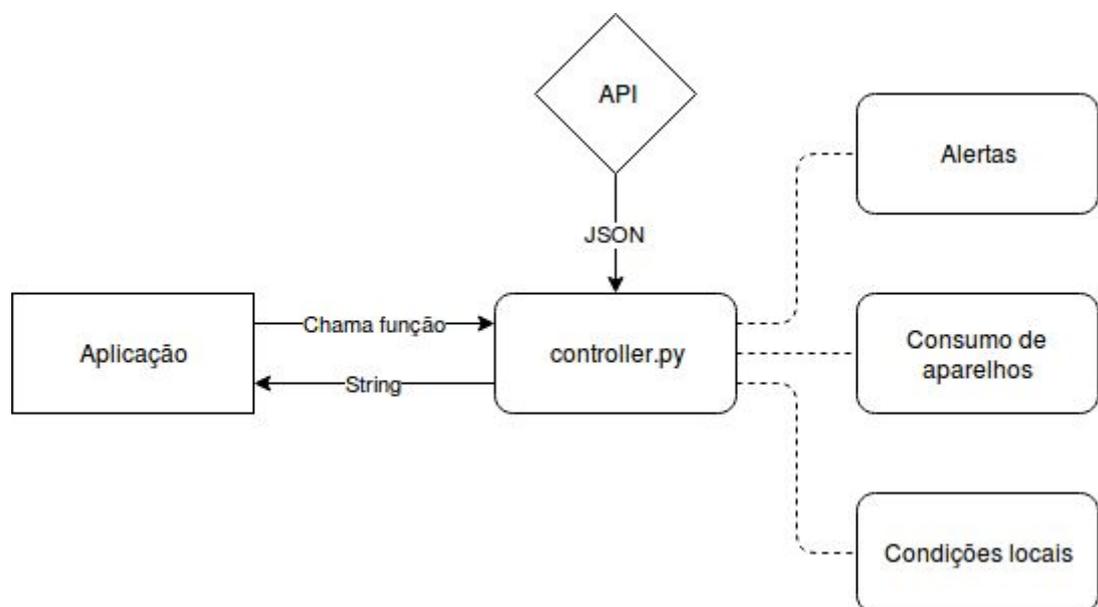


Figura 3. Funcionamento geral do módulo.

Além do processamento da resposta para retorno de strings, foi também implementado quando possível, a criação de arquivos de imagem .png com gráficos e plotagens representativas dos dados obtidos, facilitando ainda mais a assimilação das informações por parte dos usuários.

5 CONCLUSÃO

Como estagiário pela LCS Link Engenharia, tive a oportunidade de trabalhar em projetos com aplicabilidade real, adquirindo experiência a respeito do funcionamento do mercado, entendendo melhor os ciclos de desenvolvimento de software e a logística interna de um ambiente de desenvolvimento real.

Através dos projetos desenvolvidos pude ter contato com novas tecnologias, desenvolver minhas habilidades analíticas e aprofundar o entendimento que possuía a respeito dos conceitos e aplicações da linguagem Python, a qual foi utilizada primariamente ao longo do meu período de atuação dentro da organização.

A respeito dos projetos desenvolvidos, tanto os módulos de detecção de outliers quanto os de processamento de requisições encontram-se em repositórios mantidos pela equipe, sendo utilizados ao longo de outros algoritmos para fins de prototipação, a instalação do ambiente colaborativo encontra-se ativa e em uso durante as atividades de desenvolvimento.

Acredito ter alcançado meu objetivo como estagiário, realizando as atividades propostas, respeitando prazos e horários, auxiliando no progresso geral da equipe, e atuando de forma a melhor atender os interesses e objetivos de cada projeto.

Considero esta experiência de grande valor para minha formação na área, aliada aos conhecimentos adquiridos ao longo de todo o curso de Sistemas de Informação, permitindo o alcance de uma base sólida para ingresso no mercado de trabalho no futuro.

CRONOGRAMA

	2019						
	Jan	Fev	Mar	Abr	Mai	Jun	Jul
Atividades de estágio							
Escrita do relatório							
Apresentação							
Correção/Ajustes							

BIBLIOGRAFIA

- Mancini, Monica. (2018). Internet das Coisas: História, Conceitos, Aplicações e Desafios.
- Simon, Annina & Singh Deo, Mahima & Selvam, Venkatesan & Babu, Ramesh. (2016). An Overview of Machine Learning and its Applications. International Journal of Electrical Sciences & Engineering. Volume. 22-24.
- Youguo Li, Haiyan Wu. (2012). A Clustering Method Based on K-Means Algorithm. Physics Procedia. Volume 25 (1104-1109)
- J. Hodge, Victoria & Jim, Austin. (2004). A Survey of Outlier Detection Methodologies. Artificial Intelligence Review. Volume 22 (85-126)
- V. Barnet & T. Lewis. (1994). Outliers in statistical data. International Journal of Forecasting, Volume 12 (176-176)
- OpenEDG Python Institute, 2019. What is Python. Disponível em: <<https://pythoninstitute.org/what-is-python/>>. Acesso em: 07/maio/2019.