

SERVIÇO NACIONAL DE APRENDIZAGEM INDUSTRIAL  
FACULDADE DE TECNOLOGIA SENAI CAMPO GRANDE  
CURSO SUPERIOR DE TECNOLOGIA EM AUTOMAÇÃO INDUSTRIAL

LENNON PEREIRA RANGEL

BIG DATA ANALYTICS: UMA APLICAÇÃO DE COLETA DE DADOS DE UM  
SENSOR DE TEMPERATURA COM ARDUINO E MATLAB

CAMPO GRANDE/MS

2019

LENNON PEREIRA RANGEL

BIG DATA ANALYTICS: UMA APLICAÇÃO DE COLETA DE DADOS DE UM  
SENSOR DE TEMPERATURA COM ARDUINO E MATLAB

Trabalho de Conclusão de Curso submetido ao Curso de Superior de Tecnologia em Automação Industrial, da Faculdade de Tecnologia SENAI Campo Grande, como parte das exigências para conclusão do curso.

Orientador: Prof. Dr. Pedro Eugênio Marcondes Justino Ribeiro.

CAMPO GRANDE/MS

2019

LENNON PEREIRA RANGEL

BIG DATA ANALYTICS: UMA APLICAÇÃO DE COLETA DE DADOS DE UM  
SENSOR DE TEMPERATURA COM ARDUINO E MATLAB

Trabalho de Conclusão de Curso submetido ao Curso Superior de Tecnologia em Automação Industrial, da Faculdade de Tecnologia SENAI Campo Grande, como parte das exigências para conclusão do curso, sob a orientação da Prof. Dr. Pedro Eugênio Marcondes Justino Ribeiro, o qual foi apresentado perante a Banca Examinadora abaixo identificada, tendo sido aprovado, com nota final \_\_\_\_\_ e conceito \_\_\_\_\_.

Data da aprovação: \_\_\_\_/\_\_\_\_/\_\_\_\_

Prof. Dr. Pedro Eugênio Marcondes Justino Ribeiro  
Orientador

Prof. Me. Carlos Alberto Nogueira Pereira  
Examinador

Prof. Esp. Joe Alexandro Schmitz  
Examinador

CAMPO GRANDE/MS  
2019

## DEDICATÓRIA

*Dedico este trabalho primeiramente a Deus, que nos criou e me deu forças e vontade de vencer. Seus ensinamentos me inspiraram e me deram coragem para questionar realidades, além de propor sempre um novo mundo de possibilidades. Também a minha família, que muito me apoiou e me incentivou, bem como meus professores e amigos a realizá-lo com amor e carinho.*

## **AGRADECIMENTOS**

Agradeço primeiramente a Deus, pois devo a Ele a minha vida;

A minha querida mãe Eliane e meu pai do coração Maurício que estiveram presentes a cada dia dessa caminhada e que sempre me incentivaram nos estudos com palavras de otimismo e perseverança;

A minha irmã Lohany Victória, por acreditar e torcer, sempre;

Ao meu pai Oziel e sua companheira Vera que sempre torceram por mim;

Aos meus amigos pelo companheirismo e incentivo em concluir esta etapa;

Aos colegas de sala que puderam tornar essa caminhada mais amena e prazerosa e aos professores que não mediram esforços para transmitir e garantir o melhor aprendizado para mim e meus colegas;

Ao meu amigo Francisco Aragão, pela amizade, confiança, apoio moral e fé;

Ao meu orientador Pedro Eugênio por sua dedicação, compreensão e por ter confiado na minha capacidade para chegar até aqui;

Agradeço a Faculdade de Tecnologia SENAI Campo Grande e a todos que fazem parte desta importante Instituição;

A outras pessoas e entidades que contribuíram, direta ou indiretamente, para realização desta dissertação;

Enfim, muito obrigado a todos, sem vocês não seria possível a realização desse sonho.

## EPÍGRAFE

*“Eu tentei 99 vezes e falhei, mas na centésima tentativa eu consegui, nunca desista de seus objetivos mesmo que esses pareçam impossíveis, a próxima tentativa pode ser a vitoriosa.”*

*(Albert Einstein)*

## LISTA DE FIGURAS

Figura 1 – Estimativa na geração de dados até 2025 .....	13
Figura 2 – Evolução e características das revoluções industriais .....	17
Figura 3 – Os 9 pilares da Indústria 4.0 .....	17
Figura 4 – Montagem do sistema.....	21
Figura 5 – Arduino UNO.....	22
Figura 6 – Interface de programação do MATLAB.....	22
Figura 7 – Sistema finalizado.....	22
Figura 8 – Leitura de temperatura.....	23
Figura 9 – Plotagem da temperatura versus o tempo .....	23
Figura 10 – Taxa de aquisição dos dados .....	23
Figura 11 – O menor incremento mensurável do sensor pelo Arduino.....	23
Figura 12 – Aquisição dos dados em tempo real .....	24
Figura 13 – Plotagem dos dados gravados.....	24
Figura 14 – Suavização dos dados com filtro de média móvel.....	24
Figura 15 – Dados filtrados .....	24
Figura 16 – Arquivo com os dados de temperatura armazenados.....	25
Figura 17 – Analysis toolpack .....	25
Figura 18 – Gráfico de Histograma.....	26
Figura 19 – Cálculo da média móvel .....	26
Figura 20 – Média móvel da temperatura.....	26
Figura 21 – Média móvel do tempo .....	27
Figura 22 – Ajuste exponencial .....	27

## LISTA DE TABELAS

Tabela 1 – As três principais fases da história do <i>Big Data</i> .....	14
Tabela 2 – Setores para Aplicação de Data Analytics .....	16
Tabela 3 – Dados estatísticos .....	25
Tabela 4 – Histograma .....	26
Tabela 5 – Previsão da temperatura das últimas linhas da tabela .....	27



## LISTA DE ABREVIATURAS E SIGLAS

APPS	–	Aplicação móvel ou aplicativo móvel
ETL	–	Extract, transform, load
IA	–	Inteligência Artificial
IoT	–	Internet of Things
LED	–	Light-Emitting Diode
OLAP	–	On-Line Analytical Processing
PCs	–	Personal Computers
PWM	–	Pulse-Width Modulation
RDBMS	–	Relational Database Management System
SQL	–	Structured Query Language
USB	–	Universal Serial Bus
ZB	–	Zettabytes
TVs	–	Televisores

## SUMÁRIO

<b>1 INTRODUÇÃO.....</b>	<b>13</b>
1.1 Objetivo .....	14
1.2 Justificativa .....	14
 <b>2 REVISÃO DA LITERATURA .....</b>	<b>14</b>
2.1 Conceitos e um breve histórico.....	14
2.2 Big Data Analytics.....	14
2.2.1 Os 5 Vs do Big Data .....	15
2.3 Aplicabilidade.....	16
2.4 Indústria 4.0.....	16
2.5 Internet das Coisas (IoT) .....	17
 <b>3 SISTEMAS TRADICIONAIS VS SISTEMAS DE <i>BIG DATA</i>.....</b>	<b>18</b>
 <b>4 FERRAMENTAS DE BIG DATA ANALYTICS.....</b>	<b>19</b>
 <b>5 MATERIAIS E MÉTODOS .....</b>	<b>21</b>
5.1 Metodologia .....	21
5.2 Materiais utilizados .....	21
5.2.1 Plataforma de prototipagem – Arduino.....	21
5.2.2 Ferramenta de programação – MATLAB .....	22
 <b>6 RESULTADOS E DISCUSSÕES .....</b>	<b>22</b>
6.1 Aquisição dos dados.....	22
6.1.1 Medição de temperatura .....	22
6.1.2 Plotagem dos dados no tempo .....	23
6.1.3 Aquisição em tempo real.....	24
6.1.4 Armazenamento dos dados em um arquivo .....	25
6.2 Análise dos dados .....	25

6.2.1 Importar os dados.....	25
6.2.2 Análise Estatística.....	26
6.2.2.1 Média Móvel .....	26
6.2.3 Ajuste exponencial.....	27
<b>7 CONCLUSÕES.....</b>	<b>27</b>
<b>REFERÊNCIAS .....</b>	<b>28</b>
<b>ANEXOS.....</b>	<b>30</b>

# BIG DATA ANALYTICS: UMA APLICAÇÃO DE COLETA DE DADOS DE UM SENSOR DE TEMPERATURA COM ARDUINO E MATLAB\*

LENNON PEREIRA RANGEL\*\*

## RESUMO:

Na era digital, a quantidade de dados gerados e armazenados está se expandindo em curto período de tempo. Como resultado, o acelerado crescimento desse volume de informações criou muitos desafios. Nesta perspectiva, com o advento da Indústria 4.0, houve o surgimento e a evolução de diversas tecnologias dentre elas o *Big Data*, uma ferramenta que será muito importante para o futuro das organizações em diversos setores. Entretanto, devido ao aumento dos dados, as ferramentas de processamento e análise de dados atuais requerem alto investimento de hardware para processar o imenso volume de dados das organizações. A vista disso, com a utilização do *Big Data*, as empresas poderão tomar melhores decisões e tornarem-se mais competitivas perante seus concorrentes.

Neste artigo, foi realizado o projeto, a simulação e a execução de um sistema de *Big Data* aplicado numa planta térmica utilizando Arduino e MATLAB. Para isso, realizou-se um estudo abrangente sobre o surgimento e as características de *Big Data* e suas tendências atuais. Este artigo apresenta uma discussão sobre algumas das principais tecnologias de *Big Data* dos últimos anos. Além disso, são analisados os sistemas de processamento de dados convencionais em comparação com o *Big Data*. Este estudo também abrange técnicas de análise de dados, métodos de processamento e as oportunidades que essa técnica pode oferecer.

Para a análise do sistema utilizou-se uma das técnicas alternativas mais usadas de *Big Data*, o Excel. O processo foi determinado por apresentar uma aplicação de *Big Data* simples e de baixo custo e que pode ser facilmente implementado nas indústrias.

**PALAVRAS-CHAVE** – Big Data Analytics, Banco de Dados, Indústria 4.0, Sensor de temperatura, Arduino e MATLAB.

## ABSTRACT:

In the digital era, the quantity of generated data and storage is growing in a short period. As a result, the fast growth of this information's bulge created many challenges. In this perspective, regarding the fourth Industrial Revolution's advent, arised and peeked out many technologies, among them the *Big Data*, a tool that will be a very important for the organization's future in varied branch. However, because of the data's growth, the processing's tools and currents data's analysis need a high investment of hardware to process the huge bulge of organizations' data. By the way, with the using of *Big Data*, the companies can take the bests decisions and become more competitive before the entrants.

In this article, the project was accomplished, the simulation and the execution of a *Big data's* system applied in a thermic blueprint using Arduino and MATLAB. For this, an expansive study about the rising and the characteristics of *Big Data* and its current's trend was accomplished. This article presents a discussion about some of the principals *Big Data's* technologies in the last years. In addition, the systems of conventional data's processing are analyzed in comparative to the *Big Data*. This study also comprehends technics of data's analysis, processing methods and the opportunities that this technic can offer.

For the system's analysis one of the alternatives technics more common in the *Big Data* was used, the Excel. The process was determined by presenting a simple and low cost *Big Data's* application that can be easily implemented in industries.

**KEYWORDS** – Big Data Analytics, Databases, Industry 4.0, Temperature Sensor, Arduino and MATLAB.

---

\* Trabalho de Conclusão de Curso submetido ao Curso de Tecnologia em Automação Industrial, da Faculdade de Tecnologia SENAI Campo Grande, como parte das exigências para conclusão do curso, sob a orientação do Prof. Dr. Pedro Eugênio Marcondes Justino Ribeiro.

\*\* Acadêmico do 6º semestre do Curso Superior de Tecnologia em Automação Industrial, da Faculdade de Tecnologia SENAI Campo Grande. E-mail: lennonpereirarangel@hotmail.com.

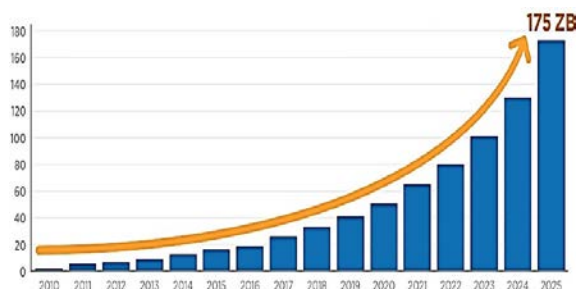
## 1 INTRODUÇÃO

O volume dos dados tem crescido de maneira descomunal nas últimas décadas. Tal crescimento se deve ao aumento do número de pessoas com acesso à tecnologia e a *Internet*, ocasionando dessa forma uma transformação no modo de processar os dados. (GALDINO, 2016).

Em razão dos atuais avanços científicos nas tecnologias de informação e comunicação e a propagação do uso de sensores e dispositivos conectados, as fontes de dados globais expandiram rapidamente. Essa produção surpreendente de dados massivos, em conjunto com as oportunidades que ele oferece para encontrar novos valores e obter novos *insights* e os diferentes desafios a qual tenta criar em termos de análise e gerenciamento, originaram um novo conceito frequentemente chamado de *Big Data*. (MACHADO, 2018).

A utilização do *Big Data* possibilitará para os diferentes tipos de indústrias e aplicações melhores tomadas de decisão. Segundo as análises da IBM (2019) Toda ação, interação, decisão, relacionamento e eventualidade que acontece em qualquer uma das infraestruturas do mundo natural e humano, hoje pode ser representada na forma de dados. Na figura 1 pode-se ver a estimativa na geração de dados até 2025.

Figura 1 – Tamanho anual da esfera de dados global.



Fonte: Adaptada de Reinsel, D.; Gantz, J.; Rydning, J. (2018).

Diante disso, tudo que está conectado à *Internet* poderá ser futuramente uma fonte de dados a ser

tratado. Em função disso, o *Big Data* é uma importante ferramenta em crescente desenvolvimento a qual vem sendo utilizada nos últimos anos para processar esse imenso volume de dados.

Nesse cenário, conforme o mundo se torna mais conectado, não somente se enfrenta uma enorme multiplicação dos dados criados pelo ser humano, mas também uma quantidade extraordinariamente crescente de dados criados por máquinas. Dessa forma, armazenar essa grande massa de dados é o grande desafio do *Big Data*, visto que não é algo barato. (MACHADO, 2018).

No entanto, com boas infraestruturas de gestão específicas e técnicas avançadas de análise de dados, essas gigantescas quantidades de dados podem ser coletadas com eficiência para otimizar as operações industriais e oferecer uma qualidade de produção mais elevada, bem como operações mais produtivas e maiores lucros.

No momento atual, é evidente que o mundo está à frente de uma nova revolução tecnológica que já vem sendo marcada pelo surgimento de grandes tecnologias. Com o *Big Data* no centro desta revolução, os dados gerados podem ser analisados de maneira eficiente.

As contribuições deste trabalho são:

- (a) Uma visão geral da história e conceitos de *Big Data* e suas tendências atuais;
- (b) Uma discussão dos sistemas tradicionais de processamento de dados em comparação com o *Big Data*.
- (c) Uma discussão das tecnologias e métodos de processamento de grandes dados;
- (d) Um estudo das oportunidades trazidas pelo *Big Data*, como também sua importância e vantagens frente aos interesses competitivos de mercado.
- (e) Desenvolvimento de um banco de dados e a análise dos dados coletados.

Essas contribuições estão presentes nas páginas de 9 a 17 e a conclusão na página 23.

## 1.1 Objetivo

O presente artigo tem como objetivos: desenvolver um sistema de coleta e análise de dados tendo em vista o esclarecimento e o funcionamento do *Big Data* em sua essência; mostrar uma aplicação prática de *Big Data* tendo em vista sua implementação num processo produtivo em uma indústria.

## 1.2 Justificativa

Em virtude do *Big Data* ser uma temática atual, verifica-se em artigos vigentes a apresentação do assunto de maneira conceitual e constantemente voltada para área empresarial. À vista disso, observou-se que seria realizável a elaboração do trabalho de pesquisa relacionado ao tema.

A aplicação prática do *Big Data* é um tópico relevante a ser exposto, uma vez que possibilita uma maior compreensão dessa ferramenta e como ela pode ser utilizada nas indústrias em diversos setores e processos.

## 2 REVISÃO DA LITERATURA

### 2.1 Conceitos e um breve histórico

Em tradução livre *Big Data* significa “grandes dados” (big: enorme, volumoso). E na década de 2000 um analista da Gartner Group cunhou o termo como é conhecido atualmente. A expressão *Big Data* é comumente usada para descrever um enorme conjunto de dados que cresce exponencialmente com o tempo. Em suma, esses dados são volumosos e complexos de tal maneira que os métodos de tratamento tradicionais não são capazes de gerenciá-los com eficiência. (MACHADO, 2018).

Apesar da definição do *Big Data* ser um tanto nova, suas origens datam desde os anos 60, no tempo em que se iniciou o surgimento dos dados, como também os

centros de tratamentos e os bancos de dados. (ORACLE, 2019).

Na tabela 1 verifica-se as 3 fases do *Big Data* e suas principais características:

Tabela 1 – As três principais fases da história do *Big Data*.

PERÍODO: 1970-2000	PERÍODO: 2000-2010	PERÍODO: 2000- Atualmente
Baseado em DBMS, conteúdo estruturado:	Conteúdo não-estruturado baseado na Web	Conteúdo para celular e baseado em sensor
<ul style="list-style-type: none"> <li>• RDBMS &amp; armazenamento de dados</li> <li>• Extrair, transformar, carregar</li> <li>• Processamento analítico online</li> <li>• Painel de controle &amp; tabela de desempenho</li> <li>• Mineração de dados &amp; análise estatística</li> </ul>	<ul style="list-style-type: none"> <li>• Recuperação e extração de informação</li> <li>• Mineração de opinião</li> <li>• Análise da Web e inteligência da Web</li> <li>• Análise de mídia social</li> <li>• Análise de redes sociais</li> </ul>	<ul style="list-style-type: none"> <li>• Análise com reconhecimento local</li> <li>• Análise centrada na pessoa</li> <li>• Análise relevante ao contexto</li> <li>• Visualização móvel</li> </ul>

Fonte: Adaptado de Big Data Framework (2019).

Pode-se citar também no período atual a Interação humano-computador, cujos eventos de comunicação entre pessoas e sistemas computacionais, se desenvolveram bastante nos últimos anos.

### 2.2 Big Data Analytics

A técnica conhecida como *Big Data Analytics* pode ser compreendida como um processo pelo qual se analisa uma grande massa de dados, com o fim de se obter resultados e informações, tais como parâmetros ocultos, correlações desconhecidas, entre outros, que podem auxiliar as empresas a tomarem melhores decisões.

Com as aplicações de Big Data Analytics é possível relatar, identificar, prever e inclusive aconselhar que ações podem ser exercidas. Com essa ferramenta é possível extrair, organizar, examinar e tratar os dados, para finalmente torná-los em informação útil para os negócios das empresas. (MACHADO, 2018).

Dessa forma, ao analisar os dados, as empresas passam a estar mais preparadas na tomada de decisões estratégicas. Com isso, pode-se alcançar uma melhor eficiência operacional e consequentemente maiores lucros. Desse modo, os dados podem ser analisados com respostas quase que a cada instante.

### 2.2.1 Os 5 Vs do Big Data

O *Big Data* tem como proposta fornecer uma abordagem consistente no tratamento dos dados. Dessa forma, o opera com base em alguns pilares importantes, o qual são divididos em 5 V's, que ajudam a entender a sua importância num processo produtivo.

- Volume

No momento presente os dados já são volumosos e o seu crescimento é inevitável. São enormes números de dados gerados a todo instante.

O volume é associado à imensa abundância de dados, oriundo das distintas atividades organizacionais, tais como: redes sociais, dados financeiros, informações de máquinas ou de dados sensoriais, dentre outros. (MACHADO, 2018).

Surpreendentemente, os PCs mostraram um enorme avanço nos últimos tempos, de modo que hoje é possível visualizar como o volume de informação aumentou em dimensões inimagináveis. Dessa forma, a tendência é que se desenvolva a todo momento. (STEVAN; LEME e SANTOS, 2018).

- Velocidade

Em consequência à difusão tecnológica, os dados acabaram sendo disseminados com maior velocidade. Diante disso, à medida em que são gerados, esses dados são examinados e tratados a cada segundo. (MACHADO, 2018).

- Variedade

A variedade se refere aos tipos de formatos do qual derivam os dados, sendo estes resultados de inúmeras fontes, tais como: e-mails, documentos, transações bancárias, dentre outros. (MACHADO, 2018). Por esse motivo, devido à variedade desses dados não há um parâmetro comum e nem sequer as mesmas características, o que indica que essas informações jamais poderiam ser tratadas em um *database* (banco de dados) convencional. Nessa situação, a presença humana por vezes faz-se necessária para analisar e compreender esses dados.

- Veracidade

Em meio a difusão dos dados, faz-se necessário analisar quais dados são verídicos e se estes estão atualizados. Uma vez desatualizados, passam a ser considerados não-verídicos visto que podem induzir à tomada de decisões incorretas. (MACHADO, 2018).

O objetivo da análise verídica é orientar de modo seguro as organizações. Por esse motivo, é importante que as empresas estabeleçam projetos que apresentem o modo como esses dados serão empregados.

- Valor

Os dados atualmente são uma das maiores riquezas de uma empresa. O valor é o tópico que mais se sobressai em comparação aos demais pilares do *Big Data*. Sem ele, nenhuma das ideias e concepções já apresentadas teriam significado, uma vez que não seria possível extrair valor das informações e além disso, analisa-las, levando em conta as atividades das organizações. (MACHADO, 2018).

Entretanto, nem todos esses dados são significantes ou benéficos para os propósitos das empresas. É preciso portanto, realizar um bom tratamento de modo que se agregue valor, tendo em vista os interesses competitivos de mercado.

## 2.3 Aplicabilidade

Existem diversos exemplos que mostram a aplicabilidade do *Big Data* nas empresas, o qual são classificados em três categorias:

- **Dados Sociais:** são dados que se originam a partir das ações comportamentais das pessoas. Como exemplos: pesquisas realizadas no Google, comentários e reações nas redes sociais, entre outros. (SANTOS, 2018).
- **Dados Corporativos:** são dados gerados pelas organizações, comumente entre os departamentos corporativos e/ou regiões geográficas. E ajudam a mensurar a produtividade e detectar gargalos. (SANTOS, 2018).
- **Dados Pessoais ou Dados de Coisas:** são dados oriundos de dispositivos ligados à rede, como por exemplo: TVs, automóveis, eletrodomésticos e outros. (SANTOS, 2018).

Na área industrial, mais precisamente na manutenção preditiva o *Big Data* tem como função por meio da análise de dados descobrir padrões considerando a probabilidade de um evento futuro ocorrer, sendo este bom ou ruim. (MACHADO, 2018). Nesse sentido, o *Big Data* é utilizado com o intuito de prever um problema antes que o mesmo ocorra. Assim, é possível fazer a correção necessária no momento exato, reduzindo custos e otimizando tempo de parada em vez de esperar que o problema aconteça, para depois encontrar uma solução.

Contudo, a técnica também pode ser utilizada na indústria em diferentes propósitos, desde a gestão até o controle e monitoramento de um processo produtivo.

Tendo em vista as funcionalidades do *Big Data*, é evidente que essa tecnologia promissora é o que divide uma empresa comum daquela que se destaca no mercado. Assim, as empresas comprometidas com o *Big Data* tendem a escolher melhor os seus

colaboradores, a uma análise de dados mais concisa e um relacionamento mais satisfatório com o cliente.

Na tabela 2 são apresentadas algumas das aplicações em diferentes setores:

Tabela 2 – Setores para Aplicação de Data Analytics.

Tipo de negócio/setor	Exemplo dos tipos de dados de entrada	Oportunidades de negócio
1 - Bancos/crédito e Seguradoras	Histórico de transações; Ficha cadastral; Proteção do crédito (SPC).	Aprovação de crédito; Análise de mercado; Detecção de fraudes.
2 - Segurança	Histórico de visitação; Ficha cadastral; Textos de notícias e de conteúdo da WEB.	Detecção de padrões de comportamentos físico ou digital que oferecem algum tipo de risco.
3 - Saúde	Prontuário médico; Sequenciamento de genomas.	Diagnóstico preditivo (previsão); Análise de dados genéticos.
4 - Óleo, gás e eletricidade	Dados de sensores distribuídos.	Otimização dos recursos de produção Predição/detecção de falhas e fraudes.
5 - Varejo	Histórico de transações; Percorso de compra em loja física e/ou virtual.	Aumento de faturamento pela otimização do mix de produto com base no padrão de comportamento durante a compra.

Fonte: HOPPEN, J.; BAPTISTA, M. (2015.)

No varejo, pode-se destacar que por meio do *feedback* dos clientes, ocorre um aumento em sua satisfação ou fidelização com a empresa.

Na saúde, uma das maiores oportunidades que podem ser alcançadas através do *Big Data* seria a descoberta de doenças e de tratamentos.

Além dessas, existem outras gamas de aplicações nos mais variados setores para diferentes fins, gerando dessa forma, novas possibilidades de negócio, melhores tomadas de decisão e consequentemente maiores lucros para as organizações.

## 2.4 Indústria 4.0

A Indústria 4.0 é um conceito disseminado pelo governo da Alemanha, além disso, pode ser



determinado como *Smart Factory* (Fábricas Inteligentes) em países como os Estados Unidos, onde o termo é empregado com o sentido de unificar os meios de produção e a tecnologia. (STEVAN; LEME; SANTOS, 2018).

Também conhecida como a 4ª Revolução Industrial, assim como as três primeiras revoluções da manufatura mundial, é marcada pelo conjunto de mudanças nos processos de fabricação, das operações e dos sistemas relacionados à produção.

Na indústria 4.0 os mundos virtuais e físicos se unem por meio da *Internet*. Ou seja, os fornecedores, distribuidores, unidades fabris e até mesmo o produto, são conectados digitalmente, possibilitando uma cadeia de valor altamente integrada. (STEVAN; LEME; SANTOS, 2018).

Na figura 2 as características das revoluções industriais anteriores até a atual:

Figura 2 – Evolução e características das revoluções industriais.



Fonte: Suporte Geográfico (2017).

A Indústria 4.0 se baseia na conectividade das fábricas com máquinas e sistemas, no qual dispõe autonomia e capacidade em agendar manutenções, com o intuito de prever falhas nos processos produtivos e se adaptarem a mudanças inesperadas que ocorrem nas linhas de produção. (MACHADO, 2018).

Stevan, Leme e Santos (2018) mencionam dados que comprovam a ideia de que uma enorme quantidade de sensores introduzidos no chão de fábrica pode gerar uma economia de até 12% em manutenção preditiva se

comparado aos reparos programados, o que acarretaria numa redução de 30% nos gastos com manutenção e 70% no período de inatividade das máquinas devido a falhas de equipamentos.

Nesse sentido, o *Big Data* sendo um dos pilares da indústria 4.0, terá um papel fundamental nesse cenário das grandes revoluções industriais, identificando falhas nos processos de fabricação, ajudando a otimizar a qualidade da produção, economizando energia e tornando mais eficiente a utilização das etapas e de recursos na produção.

Na figura 3 estão representados os 9 pilares da Indústria 4.0:

Figura 3 – Os 9 pilares da Indústria 4.0.



Fonte: Endeavor Brasil (2017).

Conclui-se portanto que, na Indústria 4.0 o *Big Data* desempenha o resultado que se espera com o progresso da revolução, promovendo assim, uma forma diferente de se trabalhar as tomadas de decisão, tendo em vista a melhoria, em eventos como a análise da produção e controle dos processos, através das ferramentas de *Big Data*, por exemplo.

## 2.5 Internet das Coisas (IoT)

O termo IoT traduzido do inglês, significa “Internet das Coisas” e, diz respeito a tudo que está conectado à *Internet*, tendo como finalidade o compartilhamento de informações, por meio de ferramentas como apps de IoT, equipamentos em conexão, maquinário

industrial, dentre outros. Os dispositivos *online* utilizam sensores para extrair dados e, em determinadas situações podem atuar neles. (MACHADO, 2018).

O primeiro dispositivo do mundo de Internet das Coisas foi criado por John Romkey, no qual posteriormente apresentou sua inovação na conferencia INTEROP, em 1989, a pedido do presidente da empresa Dan Lynch. Em sua exposição ele apresentou uma torradeira comum que podia ser ligada e desligada através da *Internet*. A partir disso, o concepção de IoT cresceu em âmbito global e não possui indícios de fim. (HARWOOD, 2019).

Assim como o *Big Data* a IoT também é um dos pilares da Indústria 4.0 e tem como propósito interconectar todos as máquinas que constituem um fábrica, reordenando por completo as suas aplicações. (MACHADO, 2018). Dessa forma os gestores podem prever a mínima indicação de perda de rendimento da produção, bem como as falhas nos processos manufatureiros. Com isso, gera-se muito mais eficiência, visto que se amplia as possibilidades de análises computacionais, a qual proporcionam suporte para as decisões.

A IoT é um dos contribuintes fundamentais do *Big Data*. Adquirir valor na Internet das Coisas, demanda análise de *Big Data*. Os métodos associados envolvem a análise preditiva, mineração de dados, serviços em nuvem, dentre outros. (MACHADO, 2018). A maior parte das empresas estão usando a associação dessas ferramentas para se conseguir maior valor da IoT.

### 3 SISTEMAS TRADICIONAIS VS SISTEMAS DE BIG DATA

Há anos a gestão dos dados é uma ideia empregada nas organizações, a qual planejam otimizar os seus sistemas por meio da administração relacionada a extração e tratamento dos dados. A distinção no tratamento de dados dos padrões convencionais (SQL

- Linguagem de Consulta Estruturada) para os padrões de *Big Data Analytics*, surge da diferença entre escalabilidade vertical e horizontal. (GALDINO, 2016).

A escalabilidade vertical é aplicada nos sistemas SQL e tem por efeito alcançar um maior processamento. Assim sendo, há um investimento em *hardwares* mais poderosos, como também em memórias, CPU, entre outros recursos. Resumindo, o objetivo é proporcionar uma atualização no servidor. Já o *Big Data* usa a escalabilidade horizontal, ou seja, adota-se sistemas de *commodities* para tarefas que antes seria inevitável o uso de supercomputadores. Dessa forma, ao dedicar várias máquinas para uma tarefa comum, a tolerância a falhas da aplicação aumenta. (VERADATA, s.d.). Logo, é possível realizar de maneira organizada, o tratamento de um grupo de dados que apenas uma única máquina não seria eficiente no processamento dos dados.

Nos padrões convencionais, o *Business Intelligence*, é uma técnica de gestão, que visa a administração de um sistema ou processo dirigido pela análise de dados, tendo como finalidade entender fatores que alteram de forma positiva ou negativa a empresa. Dessa forma, é um ótimo auxiliador em tomadas de decisões. (GALDINO, 2016).

Na maior parte dos casos, o *Business Intelligence* restringe-se em analisar o histórico dos dados, ou seja, o seu passado, a qual refere-se a anos, meses, ou dias, no entanto, sendo muito improvável ser do dia anterior. Em contra partida, com o *Big Data*, as informações podem ser bastante recentes (em se tratando de segundos, minutos ou horas), em certas ocasiões até em tempo real. Na esfera de *Big Data*, os dados tronam-se rapidamente ultrapassados e portanto inúteis. (MACHADO, 2018).

O ETL (Extração, Transformação e Carregamento), é uma ferramenta que segue o mesmo fundamento do Business Intelligence, sendo uma técnica usada em muitas empresas, tendo como função, realizar a coleta

de dados de todos os tipos e formatos, para em seguida transformá-los, por meio de algoritmos, empregando princípios de correlações entre si e carregando-os em uma interface de visualização, no qual os administradores da alta gerência, conseguirão visualizá-los, possibilitando, desse modo, a extração de informações que os permitirão tomar indicadores de melhoria nos processos empresariais. (GALDINO, 2016).

Em análise de dados, OLAP que traduzido do inglês quer dizer (Processamento Analítico *On-Line*), trata-se de um tipo de ferramenta computacional que possibilita aos usuários processar informações de forma simples, com o intuito de analisá-los por meio de diferentes concepções. Em geral, a OLAP é usada em domínios como a análise de tendências, relatórios bancários, orçamento, pressuposto das vendas, entre outros propósitos de planejamento. (ROUSE, s.d.).

Contudo, sempre que o OLAP é implementado no mundo do *Big Data*, ele não apresenta nenhum efeito, já que não é capaz de trabalhar com o aumento excessivo do volume de dados e a vasta diversidade de fontes de dados.

O RDBMS convencional (Sistema de Gerenciamento de Banco de Dados Relacional) converteu-se em um modelo bem sucedido para a administração de banco de dados no decorrer da era da *Internet*. No RDBMS as informações são estruturadas conforme o padrão relacional. (MACHADO, 2018). Contudo, hoje essa ferramenta é marcada pelo crescente enfraquecimento. A necessidade de os dados estarem devidamente estruturados faz com que haja um volume imenso de *hardware* nas empresas.

Por esse motivo, é ocasionada uma redução na performance na mesma proporção em que se aumenta o tamanho. Consequentemente, o RDBMS normalmente não é considerada uma solução para satisfazer às soluções de *Big Data*.

Resumindo, os sistemas convencionais de tratamento de dados são ótimos para processar dados estruturados, porém, suas ferramentas não são

adequadas para o tratamento de dados não estruturados.

O NoSQL, ou também “*Not Only SQL*” que traduzido do inglês seria “Não Somente SQL”, é uma estrutura totalmente distinta de bancos de dados que possibilita um processamento eficiente e alto desempenho dos dados em grandes proporções. Esse sistema, representa um suporte de banco de dados que se adaptou às grandes necessidades do *Big Data*. Por trabalhar com bancos de dados não estruturados e possuir arquitetura escalável horizontalmente, o NoSQL mostra-se mais eficiente em relação aos bancos de dados tradicionais. (MACHADO, 2018).

#### 4 FERRAMENTAS DE BIG DATA ANALYTICS

Existem várias ferramentas de *Big Data* que servem para ajudar as organizações a realizarem uma análise de dados mais eficiente e mais exata nos processos de tomada de decisão. Abaixo, algumas das principais ferramentas de *Big Data* e suas aplicações:

- Apache Hadoop

O Hadoop é uma plataforma de código aberto que muda sobretudo a forma como as empresas armazenam, processam e analisam os seus dados. A ferramenta possibilita que diferentes tipos de tarefas analíticas sejam executadas nos mesmos dados, simultaneamente e em grandes proporções no *hardware* padrão em função. (MACHADO, 2018).

O Apache Hadoop é uma das ferramentas mais importantes de *Big Data* dispostas no mercado. O *software* apresenta como uma de suas funções o aumento/redução do tamanho de arquivos em qualquer formato e totalmente de forma rápida. (ABEL, 2018).

- Import.io

O Import.io é uma plataforma utilizada para extrair dados de código aberto (*open source*), sem a

necessidade de digitar um código de acesso. Ou seja, todo o universo *web* é visto como um grande banco de dados. (ABEL, 2018).

Essa ferramenta é completa e extremamente simples de ser utilizada, não exige uma programação complementar e com apenas alguns cliques consegue-se extrair totalmente as informações de um *site* em formato de relatório completo, que posteriormente poderá ser analisado por outros *softwares*. (Equipe TD, 2018).

Para extrair os dados de uma página, os usuários adicionam uma URL e o Import.io, automaticamente, fará uma varredura no *website* e tentará extrair todas as informações consideradas mais relevantes. Além disso, é possível extrair ao mesmo tempo dados de múltiplas fontes. (ABEL, 2018).

- Machine Learning

O *Machine Learning*, traduzido do inglês (Aprendizado de Máquina) é uma técnica de análise de dados que automatiza a elaboração de modelos analíticos. É uma subdivisão da IA fundamentado na concepção de que os sistemas podem adquirir conhecimentos através dos dados, detectar padrões e com a mínima intervenção humana, tomar decisões mais assertivas. (MACHADO, 2018).

Dessa forma, no universo do *Big Data*, o *Machine Learning* é aplicado para conduzir o fluxo de informação em constante avanço e em constante transformação e gerar *insights* (compreensão, conhecimento, discernimento,) relevantes.

- Data Mining

*Data Mining*, traduzido do inglês como (Mineração de Dados), é uma das técnicas de análise de *Big Data* a qual consiste no conceito de “peneirar” os dados mais pertinentes entre todo o volume de informação coletado. (MACHADO, 2018).

Geralmente um número grande de informações não é um sinônimo de importância. Muitas das vezes, sempre haverá uma informação que pode/deve ser descartada. E essa é a função desta ferramenta. Portanto, com o *Data Mining* é possível realizar essa separação entre o que é e o que não é efetivamente importante.

Dentre as funcionalidades do *Data Mining* podem ser citadas a exploração de padrões, análises preditivas e o aprimoramento dos dados, bem como a identificação do comportamento dos consumidores, sendo capaz de descrever, precisamente, os seus perfis. (Equipe TD, 2018).

- Tableau

Uma das principais ferramentas de *Big Data* com a finalidade de visualização das informações é o Tableau. (ABEL, 2018).

O *software* permite a elaboração de mapas, inúmeros tipos de gráficos, dentre outras aplicações. Eles são gerados de forma rápida e atualizam os dados em tempo real. Existe uma versão do programa que é pública e completamente gratuita e outras que são pagas, no entanto com funções extras. (Equipe TD, 2018).

- Excel

O Excel, é uma das ferramentas alternativas de análise mais usadas no mundo, devido a sua enorme base instalada. O programa possibilita uma ótima forma de começar a aprender os conceitos básicos de análise de dados, uma vez que oferece recursos importantes, como manipular, filtrar e visualizar dados, como também plotar gráficos.

Entretanto, o Excel não é um *software* apropriado para a análise de um grande volume de dados. Isso se deve a que o número máximo de linhas em uma planilha é de mais ou menos um milhão. Em contra

partida as linhas de *Big Data* podem chegar a milhões, bilhões e até mais. Nesse ponto de vista, o Excel aparenta ser ineficiente na análise dos dados. Porém, não é o que acontece, muitas das vezes não é preciso analisar todo o banco de dados, pode-se usar apenas amostras dos dados e obter conclusões a partir delas. (DE ROSA, 2016).

Portanto, não existe a melhor ferramenta de *Big Data*, cada uma delas dependem das necessidades das empresas. Como exemplo, para algumas delas, o uso de ferramentas convencionais SQL em sistemas preventivos, que comparam vendas em diferentes épocas do ano, para planejar promoções, já são satisfatórios para os seus negócios. (GALDINO, 2016).

Logo, em todo caso, deve-se levar em conta as necessidades de negócio para então implementar a ferramenta adequada tendo em vista o resultado que se espera.

## 5 MATERIAIS E MÉTODOS

### 5.1 Metodologia

A metodologia de trabalho utilizada foi um estudo de caso a respeito do tema *Big Data*, no qual primeiramente foi realizado o estudo e posteriormente a pesquisa de referencial teórico, sendo abordados assuntos acerca das características dessa técnica. Completada esta etapa, iniciou-se o estudo de uma aplicação simples que englobasse todo um processo de *Big Data*, desde a coleta até a análise dos dados.

Em seguida, foi feito o estudo da Linguagem MATLAB, a qual foi utilizada para o desenvolvimento da programação no Arduino. Depois, foi feita a especificação dos componentes do sistema. Após a especificação dos componentes iniciou-se o processo de execução do código no Arduino.

Para o desenvolvimento deste trabalho foram utilizadas fontes de pesquisa, como artigos publicados, monografias, livros, sites, vídeos e *softwares*.

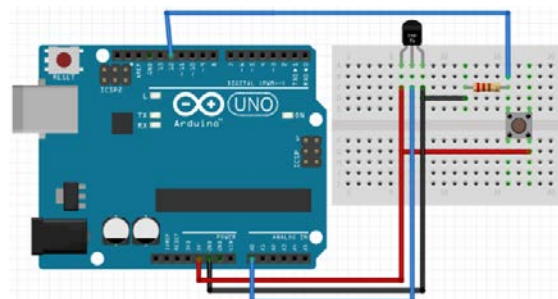
### 5.2 Materiais utilizados

Para o presente trabalho foram utilizados os seguintes materiais:

- Um Arduino modelo: UNO;
- Um Cabo USB A para B, para comunicação do computador com o Arduino;
- Uma Breadboard - mini;
- Um sensor de temperatura tipo: LM35;
- Jumpers;
- Um botão;
- Um resistor 330Ω;

Na figura 4 estão representados o esquema com as ligação do sensor de temperatura LM35 no arduino:

Figura 4 – Montagem do sistema.



Fonte: Própria.

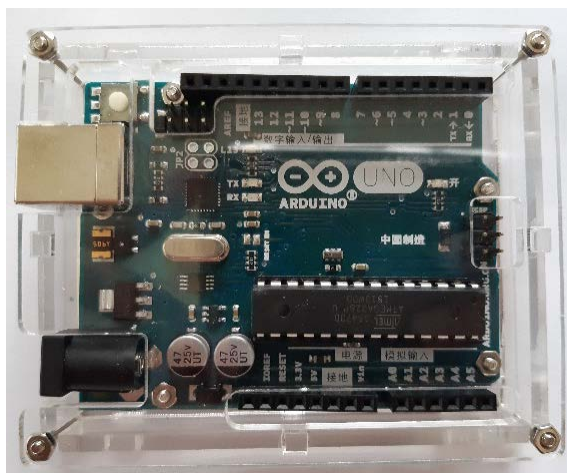
#### 5.2.1 Plataforma de prototipagem – Arduino

O arduino escolhido foi o tipo UNO, o qual possui uma placa de microcontrolador com processador ATmega328P. Dispõe de 14 pinos digitais de entrada/saída (D0-D13), (sendo que 6 destes podem ser utilizados como saídas PWM), 6 entradas analógicas (A0-A5), um conector USB, um conector de alimentação, um cabeçalho ICSP e um botão de *reset*. Na figura 5 pode-se visualizar as especificações ditas acima.

A seguir, algumas especificações técnicas sobre o Arduino UNO:

- Tensão de funcionamento: 3,3V ou 5V;
- Tensão de entrada (sugerido): 7-12V;
- Corrente CC por I/O: 40 mA;
- Corrente CC no pino 3.3V: 50 mA;

Figura 5 – Arduino UNO.



Fonte: Própria.

### 5.2.2 Ferramenta de programação – MATLAB

A ferramenta MATLAB foi utilizada para a programação do Arduino como também para a análise dos dados coletados. Esta ferramenta alternativa permite a criação, documentação e análise das aplicações de *Big Data*.

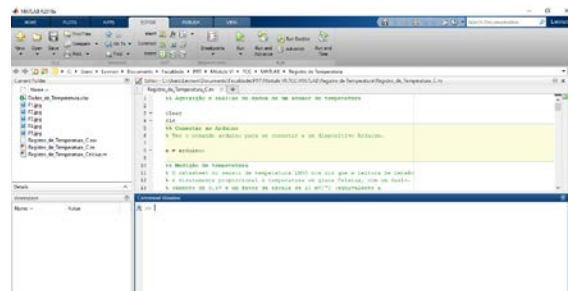
Além disso, o software oferece recursos de simulação (execução do código no computador) e de supervisão (monitoramento das variáveis, com facilidade de alterações, também pelo computador).

O MATLAB, é *software* de programação elaborado em especial para engenheiros e cientistas. O programa usa a linguagem MATLAB, a qual é fundamentada em códigos matriciais possibilitando assim uma representação mais genuína da matemática computacional.

O programa é voltado para operações numéricas, operação com matrizes, processamento de sinais e plotagem de gráficos em ambientes cujas soluções são expressas da mesma forma como são escritas matematicamente.

Na figura 6, a interface inicial de programação do MATLAB:

Figura 6 – Interface de programação do MATLAB.



Fonte: Própria.

## 6 RESULTADOS E DISCUSSÕES

Para utilizar o arduino no MATLAB deve-se inicialmente instalar os *drives* para realizar a comunicação e execução do código. Feita a conexão, basta enviar o código para o Arduino. Na figura 7, os componentes instalados na placa Arduino.

Figura 7 – Sistema finalizado.



Fonte: Própria.

Para este trabalho, não foi utilizado a IDE Arduino. Todo o código foi desenvolvido no MATLAB.

O código foi adaptado de (Madhu Govindarajan, 2016) e encontra-se em Anexos.

### 6.1 Aquisição dos dados

#### 6.1.1 Medição de temperatura

Inicialmente, foi feita a medição de temperatura do sensor em graus Celsius, que, de acordo com o



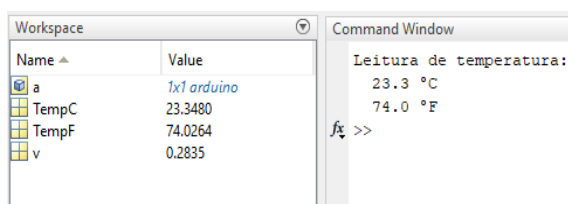
*datasheet* é proporcional a tensão de saída do pino em que está conectado à placa Arduino.

Após a conversão da tensão em graus Celsius converteu-se o resultado para Fahrenheit. Como mostra a figura 8.

Para uma medição mais precisa, inseriu-se uma constante que subtrai o valor da tensão aferida no pino de saída em que está conectado o sensor. Dessa forma, a conversão pode ser representada na equação 1:

$$\text{TempC} = (v - 0.05) * 100; \quad (1)$$

Figura 8 – Leitura de temperatura.



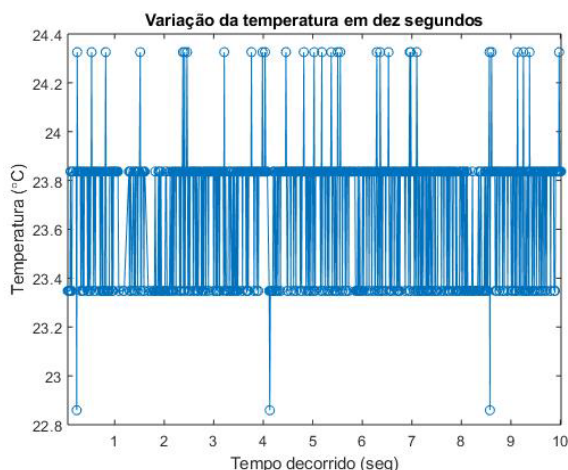
Fonte: Própria.

### 6.1.2 Plotagem dos dados no tempo

Após a leitura da temperatura, os dados foram registrados em um intervalo de 10 segundos e plotados no gráfico de temperatura (°C) em função do tempo (s). Como mostra a figura 9.

É possível visualizar também a faixa (mínima e máxima) de variação da temperatura.

Figura 9 – Plotagem da temperatura versus o tempo.

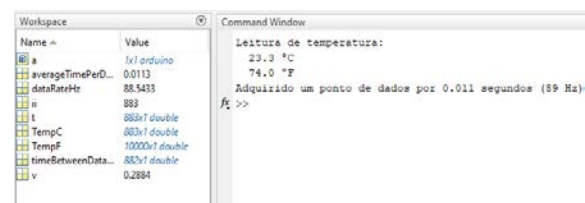


Fonte: Própria.

Depois, foi feito o cálculo da taxa de aquisição dos dados, ou seja, a cada 0.011 segundos foi adquirido um ponto de dados a uma frequência de 89 Hz. Como pode ser visto na figura 10.

Essa frequência representa o quão rápido o MATLAB pode coletar os dados.

Figura 10 – Taxa de aquisição dos dados.



Fonte: Própria.

Em seguida, foi realizado o cálculo do menor incremento mensurável do sensor pelo Arduino. Isto é, a menor tensão e temperatura (°C e °F) possível.

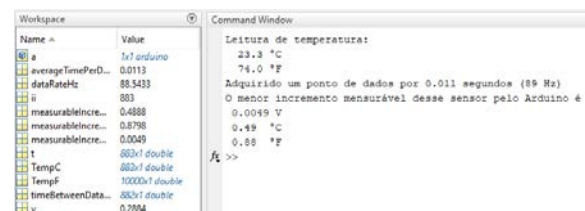
Essa função determina a velocidade mais rápida com o qual foi possível adquirir os dados.

Como pode-se observar, na figura 9 os dados são muito irregulares. Isso ocorre porque o Arduino utilizado é um dispositivo de 8 bits. E apenas lê valores entre 0 e 1023 nos pinos analógicos.

Nesse sentido, quando se usa esses valores, verifica-se que a leitura corresponde a aproximadamente 0,5 graus Celsius e 1 Fahrenheit, aproximadamente, como mostra a figura 11. Isso explica o motivo pelo qual os dados são tão instáveis.

Logo, mesmo uma pequena variação no valor da tensão significa que há uma mudança considerável na temperatura.

Figura 11 – O menor incremento mensurável do sensor pelo Arduino.



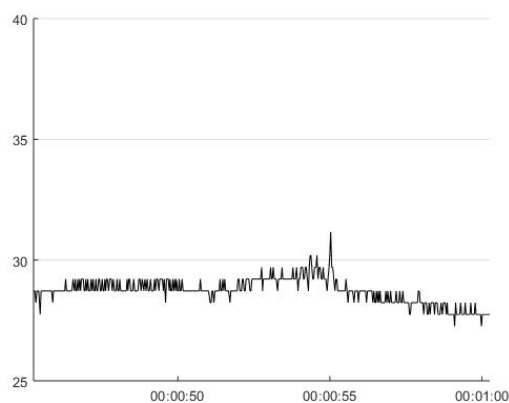
Fonte: Própria.

### 6.1.3 Aquisição em tempo real

Depois, foi utilizada uma função que representa uma linha animada, a qual simplifica o fluxo de novos dados para um gráfico, onde é possível visualizar como o valor medido muda com o tempo. Como mostra a figura 12. Um aspecto vantajoso dessa função é que os dados são armazenados automaticamente no gráfico, sem a necessidade de registrá-los.

O código possui um *script* em que ao pressionar o botão, interrompe-se a coleta dos dados.

Figura 12 – Aquisição dos dados em tempo real.

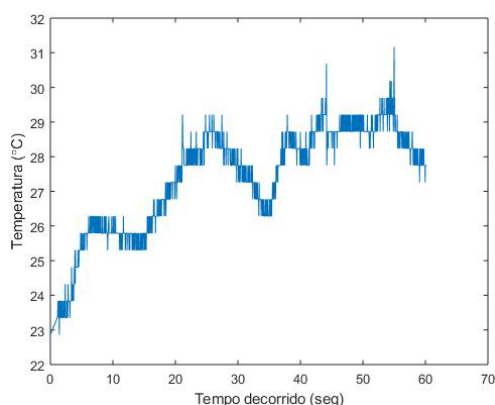


Fonte: Própria.

Na figura 13, é possível visualizar por meio de um gráfico todos os dados gravados do início ao final da aquisição, que foi de mais ou menos 1min.

Devido aos ruídos os dados precisam de algum pós-processamento.

Figura 13 – Plotagem dos dados gravados.



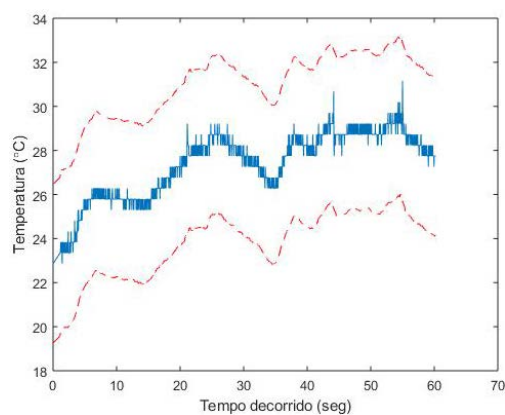
Fonte: Própria.

O ruído de alta frequência no sinal de temperatura pode ser removido aplicando-se um filtro de média móvel. O filtro de média móvel trata-se da média de um conjunto de várias amostras de temperatura.

Com base na folha de dados, sabe-se que o sensor de temperatura LM35 tem uma tolerância de mais ou menos 0.5 °C à temperatura ambiente.

Assim, usando essas informações, foi calculado o maior e o menor valor possível e plotado os dados coletados no gráfico, como mostra a figura 14.

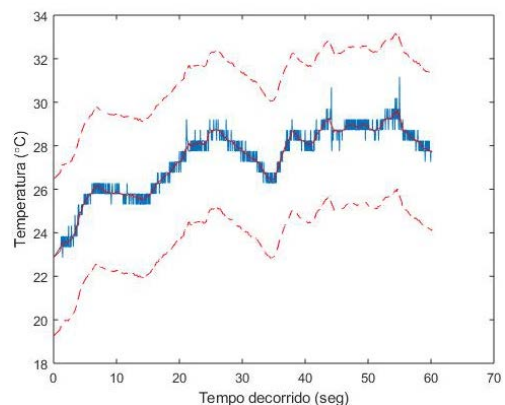
Figura 14 – Suavização dos dados com filtro de média móvel.



Fonte: Própria.

Na figura 15 a linha contínua entre a gravação dos dados representa os dados filtrados. E a tolerância do sensor de temperatura é representada pelas linhas pontilhadas na figura abaixo.

Figura 15 – Dados filtrados.



Fonte: Própria.



### 6.1.4 Armazenamento dos dados em um arquivo

Realizado todo o processo de aquisição, os dados foram salvos em um arquivo.

Para isso, foram armazenados em uma tabela, que em seguida, foi salva em formato .xlsx. contendo todas as informações necessárias, como mostra a figura 16.

Figura 16 – Arquivo com os dados de temperatura armazenados.

	A	B
	DadosdeTemperatura	
	Time_seg	Temp_C
	Number	Number
1	Time_seg	Temp_C
2	0	22.8592
3	1.1080	23.3480
4	1.1560	23.3480
5	1.2090	23.8368
6	1.2560	23.3480
7	1.3450	23.8368
8	1.3780	23.8368
9	1.4330	22.8592
10	1.4810	23.8368
11	1.5140	23.8368
12	1.5640	23.8368
13	1.5930	23.3480
14	1.6400	23.3480
15	1.6740	23.3480

Fonte: Própria.

A tabela apresenta duas colunas, sendo a primeira a do tempo em que cada temperatura foi medida e o segunda com o valor da temperatura em graus centígrados.

## 6.2 Análise dos dados

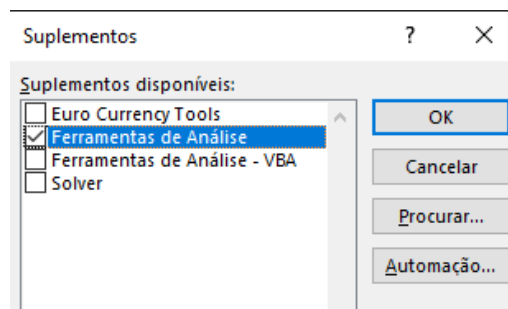
Após a aquisição, foi feita a análise dos dados, no qual foi possível descobrir tendências e correlações entre os dados.

### 6.2.1 Importar os dados

Inicialmente os dados foram executados no Excel, é foi utilizada a ferramenta de análise de dados (*analysis toolpack*) presente na aba 'dados', como mostra a figura 17.

Essa ferramenta, foi ativada nas configurações do Excel, em: (Arquivos > Opções > Suplementos > Ir...).

Figura 17 – Analysis toolpack.



Fonte: Própria.

De início, foi usada a ferramenta 'estatística descritiva' para calcular os dados estatísticos, acerca do banco de dados, como mostra a tabela 3.

Tabela 3 – Dados estatísticos.

Temperatura	
Média	27,4759353
Mediana	27,74682307
Modo	28,72434018
Mínimo	22,85923754
Máximo	31,16813294
Soma	50912,90811
Contagem	1853

Fonte: Própria.

Com base nessas informações é possível estabelecer uma temperatura ideal, por exemplo em uma máquina em operação numa indústria e, dessa forma, visualizar os melhores horários de funcionamento, bem como estipular limites de temperatura e visando uma melhor tomada de decisão, desse modo, evitando que a máquina sofra algum defeito e necessite de reparo. Já que o reparo antecipado desperdiça dinheiro e o reparo tardio provoca uma parada custosa.

Com a análise de dados do Excel é possível determinar também a frequência com os quais os valores se repetem e os que menos se repetem, com a ferramenta histograma, como mostra a tabela 4.

Tabela 4 – Histograma.

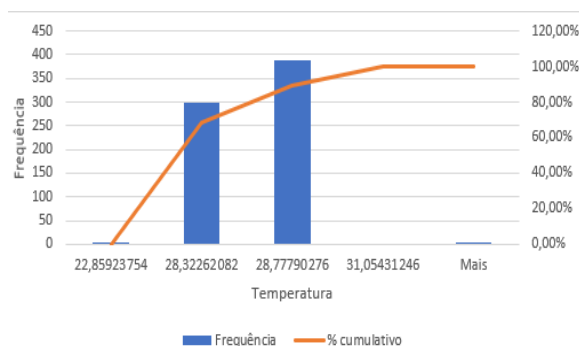
Temperatura	Frequência	% cumulativo
22,85924	1	0,05%
28,32262	299	68,41%
28,7779	387	89,31%
31,05431	0	99,95%
Mais	1	100,00%

Fonte: Própria.

Como pode ser visto na figura 18, a ferramenta de análise Histograma teve a função de calcular as frequências específicas e acumulativas de uma amostra de dados. Este método reuni dados de uma quantidade de repetições de um valor em um banco de dados.

No entanto, devido as diversas linhas de frequência, foi retirado apenas uma amostra dos valores.

Figura 18 – Gráfico de Histograma.



Fonte: Própria.

Por meio dessa análise é possível verificar por exemplo, o desempenho de uma máquina/ferramenta e até mesmo determinar sua vida útil, dentre outros fatores.

## 6.2.2 Análise Estatística

### 6.2.2.1 Média Móvel

Posteriormente, como mostra a figura 19, foi feita a média móvel de uma amostra dos dados que teve a função de estimar os próximos valores de temperatura. A média móvel proporciona informações de tendência

que uma média comum não demonstraria. (Support Office, s.d.).

A média móvel atenua os dados de temperatura de modo a elaborar um indicador de tendência linear. Seu objetivo não é prever a temperatura, mas sim, definir a sua direção atual com um pequeno atraso. (Support Office, s.d.).

Isso ocorre porque ela se baseia nos dados anteriores para calcular a média. Contudo, ela ajuda a suavizar a temperatura e a filtrar o ruído.

Ainda assim, essa simples ferramenta, pode ser muito útil na manutenção preditiva visto que se assemelha muito com a análise preditiva de dados.

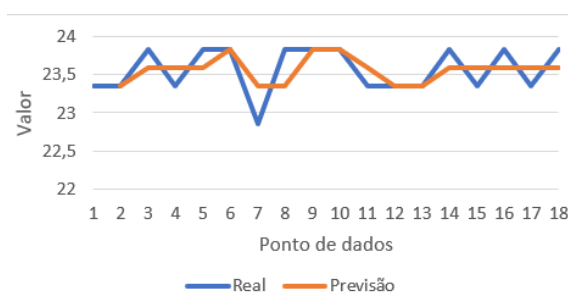
Figura 19 – Cálculo da média móvel.

	A	B	C	D	E
1	Time_seg	Temp_C		#N/D	
2	0	22,85923754		=MÉDIA(B3:B4)	
3	1,108	23,34799609		MÉDIA(núm1; [núm2]; ...)	
4	1,156	23,34799609		23,59237537	
5	1,209	23,83675464		23,59237537	
6	1,256	23,34799609		23,83675464	
7	1,345	23,83675464		23,34799609	
8	1,378	23,83675464		23,34799609	
9	1,433	22,85923754		23,83675464	
10	1,481	23,83675464		23,83675464	
11	1,514	23,83675464		23,59237537	
12	1,564	23,83675464		23,34799609	
13	1,593	23,34799609		23,34799609	
14	1,64	23,34799609		23,59237537	
15	1,674	23,34799609		23,59237537	
16	1,709	23,83675464		23,59237537	
17	1,746	23,34799609		23,59237537	
18	1,775	23,83675464		23,59237537	
19	1,814	23,34799609			
20	1,841	23,83675464			

Fonte: Própria.

Na figura 20 pode-se visualizar a média móvel aplicada no intervalo de 0 a 20 da tabela dos dados de temperatura:

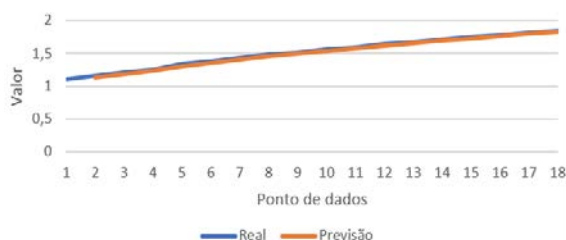
Figura 20 – Média móvel da temperatura.



Fonte: Própria.

É possível realizar a média móvel também para o tempo em que foram adquiridos os dados e estipular o período futuro. Como mostra a figura 21. Entretanto, como o tempo é constante, quase não há variações.

Figura 21 – Média móvel do tempo.



Fonte: Própria.

### 6.2.3 Ajuste exponencial

Diferentemente da média móvel, a ferramenta de ajuste exponencial, prevê os próximos valores de temperatura baseado nos dados anteriores. (Support Office, s.d.). O cálculo foi utilizado nas últimas 20 linhas da tabela. Como pode ser visto na tabela 5.

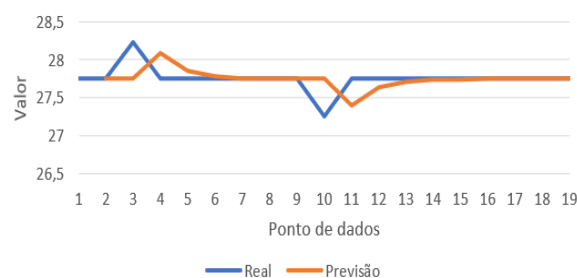
Tabela 5 – Previsão da temperatura das últimas linhas da tabela.

Previsão de temperatura
27,84946
27,77761
27,75606
27,74959
27,74765
27,74707
27,40477
27,64421
27,71604
27,73759
27,74405
27,74599
27,74657
27,74675
27,7468

Fonte: Própria.

Na figura 22 é possível visualizar o gráfico com um visão geral de como seriam as medições em relação as anteriores.

Figura 22 – Ajuste exponencial.



Fonte: Própria.

Assim como na Média Móvel, também é possível aplicar o ajuste exponencial para o tempo, para prever o período futuro. Entretanto, como o tempo é constante, as variações são mínimas.

Portanto, o método de previsão de ajuste exponencial pode ser implementado nas indústrias, uma vez que apresentaria bons resultados para previsões de curto prazo baseado nos dados históricos, para se ter uma melhor tomada de decisão.

## 7 CONCLUSÕES

Neste trabalho realizou-se a elaboração de um sistema de *Big Data* em uma planta térmica. Inicialmente foi explorado um estudo das tecnologias de *Big Data* e as diferentes técnicas utilizadas para processamento e análise de dados, bem como as vantagens do *Big Data* frente aos métodos de análise tradicionais. Foi discutido também as tendências atuais para ajudar a entender a rápida disseminação no *Big Data* e analisou-se as oportunidades e benefícios que ele fornece a sociedade como um todo.

O código implementado no MATLAB atendeu aos requisitos de simplicidade e objetividade necessários para facilitar a coleta de dados do sistema.

Para a análise dos dados, utilizou-se a ferramenta de análises de dados estatísticas do Excel, a qual mostrou-se muito eficiente para análise e tendências de dados. Esta ferramenta se mostrou eficaz e satisfatória para as aplicações da planta. Em vista disso, seria muito

eficiente aplicada numa máquina/ferramenta ou num processo produtivo.

Os resultados da análise mostraram-se promissores, atendendo por inteiro aos requisitos do trabalho. Logo, o sistema prático comportou-se de modo satisfatório, atendendo todos os objetivos propostos para a sua implementação.

É importante destacar o caráter educativo do sistema desenvolvido, visto que o mesmo pode ser usado para fins de aprendizagem na área de automação industrial.

Atualmente é notório e importante o papel do *Big Data* para as indústrias. Os grandes dados envolvem muitos desafios e benefícios significativos. Para estudos futuros, seria interessante investigar abordagens baseadas em novos dados que usam a análise de *Big Data* (análise descritiva, preditiva e prescritiva) para lidar com os problemas encontrados nas indústrias. Nesse âmbito, a análise preditiva, como um método emergente e promissor, pode ser o mais estudado, a fim de propor novas ideias de aprimoramento analítico preditivo para o apoio à tomada de decisão em aplicações industriais.

Para trabalhos futuros planeja-se projetar uma otimização nesse sistema, substituindo o sensor de temperatura LM35 por um sensor de temperatura industrial, como por exemplo o Termopar para aplicações industriais que exijam a coleta de dados em tempo real e também utilizar as ferramentas de *Big Data* como o Hadoop e a Machine Learning para realizar uma melhor coleta e análise dos dados.

Conclui-se com este trabalho que as ferramentas e técnicas atuais realizam o tratamento de dados de maneira insuficiente. O propósito de todas as técnicas analíticas de tratamento existentes é processar apenas quantidades limitadas de dados estruturados. Dessa forma, a utilização das ferramentas existentes para o processamento de *Big Data* resulta em perda de eficiência e também gera muitas complicações. Em contrapartida o *Big Data* está impactando

metodicamente a criação de valor nas organizações gerando assim, novas possibilidades de negócio, melhores tomadas de decisão e consequentemente maiores lucros.

## REFERÊNCIAS

ABEL, Carol. **Análise de dados: conheça as 8 principais ferramentas de Big Data para usar nos negócios.** Disponível

em:<<https://mindminers.com/blog/ferramentas-de-big-data/>>. Acesso em: 02 out. 2019.

BIG DATA FRAMEWORK. **Where does ‘Big Data’ come from?.** Disponível em:<

<https://www.bigdataframework.org/short-history-of-big-data/>>. Acesso em: 13 out. 2019.

DE ROSA, Miro. **How to Analyze Big Data with Excel.** Disponível

em:<<https://www.datasciencecentral.com/profiles/blogs/how-to-analyze-big-data-with-excel>>. Acesso em: 02 out. 2019.

ENDEAVOR BRASIL. **Indústria 4.0: as oportunidades de negócio de uma revolução que está em curso.** Disponível

em:<<https://endeavor.org.br/tecnologia/industria-4-0-oportunidades-de-negocio-de-uma-revolucao-que-esta-em-curso/>>. Acesso em: 18 out. 2019.

Equipe TD. **Ferramentas de Big Data: qual a melhor para sua empresa?** Disponível

em:<<https://transformacaodigital.com/ferramentas-de-big-data-qual-a-melhor-para-sua-empresa/>>.

Acesso em: 02 out. 2019.

GALDINO, Natanael. **Big Data: Ferramentas e Aplicabilidade.** In: SIMPÓSIO DE EXCELÊNCIA

EM GESTÃO E TECNOLOGIA. 13, 2016, Resende. **Anais. AEDB.**

GOVINDARAJAN, Madhu. **Plotting Live Data of a Temperature Sensor using Arduino and MATLAB.** Disponível em:<<https://la.mathworks.com/videos/plotting-live-data-of-a-temperature-sensor-using-arduino-and-matlab-121317.html>>. Acesso em: 16 out 2019.

HARWOOD, Trevor. **Internet of Things (IoT) History.** Disponível em:<<https://www.postscapes.com/iot-history/>>. Acesso em: 23 out. 2019.

HOPPEN, J.; BAPTISTA, M. **14 setores para aplicação de Data Analytics.** Disponível em:<<https://www.aquare.la/14-setores-para-aplicacao-de-data-analytics/>>. Acesso em: 29 out. 2019.

IBM. **A IBM e o uso responsável dos dados.** Disponível em:<<https://www.ibm.com/blogs/ibm-comunica/a-ibm-e-o-uso-responsavel-dos-dados/>>. Acesso em: 07 out. 2019.

MACHADO, Felipe. **Big Data: O Futuro dos Dados e Aplicações.** 1. ed. São Paulo: Érica, 2018.

ORACLE. **The Definition of Big Data.** Disponível em:<<https://www.oracle.com/big-data/guide/what-is-big-data.html>>. Acesso em: 11 out. 2019.

REINSEL, D.; GANTZ, J.; RYDNING, J. **The Digitization of the World: From Edge to Core.** Disponível em:<<https://www.seagate.com/files/www-content/our-story/trends/files/idc-seagate-data-age-whitepaper.pdf>>. Acesso em: 19 out. 2019.

ROUSE, Margaret. **OLAP (processamento analítico online).** Disponível

em:<<https://searchdatamanagement.techtarget.com/definition/OLAP>>. Acesso em: 29 out. 2019.

SANTOS, Pedro. **Big Data: o que é, para que serve e como aplicar?.** Disponível em:<<https://inteligencia.rockcontent.com/big-data/>>. Acesso em: 24 out. 2019.

STEVAN, Sergio; LEME, Murilo; SANTOS, Max. **Indústria 4.0: fundamentos, perspectivas e aplicações.** 1. ed. São Paulo: Érica, 2018.

SUPORTE GEOGRÁFICO. **INFOGRÁFICOS - REVOLUÇÕES INDUSTRIAIS.** Disponível em:<<https://suportegeografico77.blogspot.com/2017/12/infograficos-revolucoes-industriais.html>>. Acesso em: 18 out. 2019.

SUPPORT OFFICE. **Usar Ferramentas de Análise para executar análises de dados complexas.** Disponível em:<<https://support.office.com/pt-br/article/usar-ferramentas-de-an%C3%A1lise-para-executar-an%C3%A1lises-de-dados-complexas-6c67ccf0-f4a9-487c-8dec-bdb5a2cefab6?NS=EXCEL&Version=19&SysLcid=1046&UiLcid=1046&AppVer=ZXL190&HelpId=xlaaddin.chm1780&ui=pt-BR&rs=pt-BR&ad=BR>>. Acesso em: 05 nov. 2019.

VERADATA. **Escalabilidade: Conceitos e boas práticas.** Disponível em:<<http://veradata.com.br/pt-br/blog/escalabilidade-conceitos-e-boas-praticas>>. Acesso em: 28 out. 2019.

## ANEXOS

### AQUISIÇÃO DE DADOS DE UM SENSOR DE TEMPERATURA

```
clear
clc
```

#### Conectar ao Arduino

% Use o comando arduino para se conectar a um dispositivo Arduino.

```
a = arduino;
```

#### Medição de temperatura

% O datasheet do sensor de temperatura LM35 nos diz que a leitura de tensão  
% é diretamente proporcional à temperatura em graus Celsius, com um deslo-  
% camento de 0,5V e um fator de escala de 10 mV/°C (equivalente a  
% 100°C/V).  
% Portanto, a conversão pode ser representada como:  
 $T_C = (V - 0.05) * 100$   
%  
% Podemos ler a tensão de saída, convertê-la para Celsius e converter o  
% resultado para Fahrenheit da seguinte forma:

```
v = readVoltage(a,'A0');
TempC = (v-0.05)*100;
TempF = 9/5*TempC + 32;
fprintf('Leitura de
temperatura:\n %.1f °C\n %.1f °F\n',TempC,TempF)
```

#### Registre e plote 10 segundos de dados de temperatura

```
ii = 0;
TempC = zeros(1e4,1);
```

```
t = zeros(1e4,1);
tic
while toc < 10
    ii = ii + 1;
    % Ler o valor atual da tensão
    v = readVoltage(a,'A0');
    % Calcular a temperatura da tensão (com base no
    datasheet)
    TempC(ii) = (v-0.05)*100;
    TempF = 9/5*TempC + 32;
    % Obter o tempo desde o início
    t(ii) = toc;
end
```

% Pós-processamento e plotagem dos dados.  
Primeiro remova qualquer excesso  
% de zeros nas variáveis de log.  
 $TempC = TempC(1:ii);$   
 $t = t(1:ii);$   
% Temperatura da plotagem versus tempo  
figure  
plot(t,TempC,'-o')  
xlabel('Tempo decorrido (seg)')  
ylabel('Temperatura (\text{°C})')  
title('Dez segundos de dados de temperatura')  
set(gca,'xlim',[t(1) t(ii)])

#### Cálculo da taxa de aquisição

```
timeBetweenDataPoints = diff(t);
averageTimePerDataPoint =
mean(timeBetweenDataPoints);
dataRateHz = 1/averageTimePerDataPoint;
fprintf('Adquirido um ponto de dados por %.3f
segundos (%.f Hz)\n',...
averageTimePerDataPoint,dataRateHz)
```

#### Por que meus dados são tão instáveis?

```
measurableIncrementV = 5/1023;
```

```

measurableIncrementC =
measurableIncrementV*100;
measurableIncrementF = measurableIncrementC*9/5;
fprintf('O menor incremento mensurável desse sensor
pelo Arduino é\n %-6.4f V\n %-6.2f°C\n %-
6.2f°F\n',...

measurableIncrementV,measurableIncrementC,measu
rableIncrementF);

```

#### Adquira e exiba dados ao vivo

```

figure
h = animatedline;
ax = gca;
ax.YGrid = 'on';
ax.YLim = [20 40];
stop = false;
startTime = datetime('now');
while ~stop
    % Ler o valor atual da tensão
    v = readVoltage(a,'A0');
    % Calcular a temperatura da tensão (com base no
datasheet)
    TempC = (v-0.05)*100;
    TempF = 9/5*TempC + 32;
    % Obter hora atual
    t = datetime('now') - startTime;
    % Adicionar pontos à animação
    addpoints(h,datenum(t),TempC)
    % Atualizar eixos
    ax.XLim = datenum([t-seconds(15) t]);
    datetick('x','keeplimits')
    drawnow
    % Verifique a condição de parada
    stop = readDigitalPin(a,'D12');
end

```

#### Plote os dados gravados

```
[timeLogs,tempLogs] = getpoints(h);
```

```

timeSecs = (timeLogs-timeLogs(1))*24*3600;
figure
plot(timeSecs,tempLogs)
xlabel('Tempo decorrido (seg)')
ylabel('Temperatura (\circC)')

```

#### Suavizar leituras com filtro de média móvel

```

smoothedTemp = smooth(tempLogs,25);
tempMax = smoothedTemp + 2*9/5;
tempMin = smoothedTemp - 2*9/5;
figure
plot(timeSecs,tempLogs, timeSecs,tempMax,'r--
',timeSecs,tempMin,'r--')
xlabel('Tempo decorrido (seg)')
ylabel('Temperatura (\circC)')
hold on

```

#### Plote o sinal de temperatura original e suavizado e ilustre a incerteza.

```
plot(timeSecs,smoothedTemp,'r')
```

#### Salvar resultados em um arquivo

```

T =
table(timeSecs',tempLogs','VariableNames',{ 'Time_se
g','Temp_C'});
filename = 'Dados_de_Temperatura.xlsx';
% Gravar tabela em arquivo
writetable(T,filename)
% Imprimir confirmação na linha de comando
fprintf('Tabela de resultados com %g de medições de
temperatura salvas no arquivo %s\n',...

length(timeSecs),filename)

```