

# ANÁLISE E VISUALIZAÇÃO DE DADOS NA AGRICULTURA COM DATA SCIENCE

Guilherme de Cleve Farto  
*guilherme.farto@gmail.com*

João Victor Franco Rosa  
*joaovictorfrancorosa@gmail.com*

**RESUMO:** Com a crescente demanda por alimentos no mundo e a questão ambiental cada vez mais em pauta, os produtores agrícolas buscam na tecnologia formas de melhorar a produtividade das áreas agricultáveis já existentes. A análise de dados tem papel importante no crescimento da produtividade em propriedades agrícolas e agroindústrias buscando encontrar padrões para melhoria dos processos produtivos. O objetivo deste trabalho de pesquisa é analisar como a utilização da ciência de dados e a visualização dos dados são aplicados na agricultura para auxiliar técnicos ou engenheiros agrônomos e produtores rurais nas tomadas de decisão na condução da lavoura.

**PALAVRAS-CHAVE:** Agricultura; Ciência de Dados; Informação; Tecnologia;

**ABSTRACT:** With the growing demand for food in the world and the increasing concern over environmental issues, agricultural producers seek in technology ways to improve the productivity of existing agricultural areas. Data analysis has an important role in the growth of productivity in agricultural properties and agro-industries, by finding patterns for improving production processes. The objective of this work is to analyse how the use of data science and the data visualization are applied in agriculture to assist technicians, agronomists and rural producers in decision-making in growing crops.

**KEYWORDS:** Agriculture; Data Science; Information; Technology;

## 1. Introdução

O “Relatório de Recursos Mundiais: Criando um Futuro Sustentável para a Alimentação” da ONU (2019), diz que até o ano de 2050 a população mundial ultrapassará as 10 bilhões de pessoas. Com isso, a demanda por alimentos aumentar-se-á de forma exponencial, sendo necessária que a produtividade também se desenvolva cerca de 70% em relação à produção de 2010.

Além do aumento da população mundial, outro desafio para a agricultura mundial é a produção de energia a partir de fontes renováveis. A matéria prima básica para produção destas energias são as *commodities* agrícolas, ou seja, a produção de energia e a produção de alimentos devem ocorrer de forma crescente, sem que um prejudique o outro (PEREIRA, 2017).

Segundo a FAO (2019), os sensores no campo, presentes em tratores, drones e os mais diversos equipamentos, fornecem uma quantidade variada de dados sobre as qualidades químicas e físicas do solo, sobre o clima, além de infestações de pragas, o monitoramento da evolução das culturas, como taxas de crescimento, bem como diversos fatores que influenciam a produção agrícola. A análise deste montante de dados pode tornar o setor agrícola mais produtivo e eficiente, auxiliando o produtor nas tomadas de decisões como a escolha da cultivar e em quais áreas há maior necessidade de nutrientes, promovendo inovações, facilitando atividades relacionadas como seguros, financiamentos e força de trabalho.

## 2. *Smart Agriculture*

Desde o início de sua história, o Brasil é um país agrícola. Com a chegada de portugueses e espanhóis a primeira fonte de exploração foi o Pau-Brasil, onde posteriormente a sua retirada começaram os plantios de culturas domesticadas, principalmente cana-de-açúcar (SOUZA, 2017).

O agronegócio brasileiro até o início do século XX era basicamente a exploração da matéria primária, o produto saído do campo sem qualquer forma de agregação de valor, exceto o açúcar. Posteriormente, com incentivos governamentais, iniciou-se o processo de industrialização, composto basicamente por agroindústrias, cerca de  $\frac{2}{3}$ , principalmente alimentícias e têxteis (FILHO & PAULA, 2001).

Após a segunda guerra mundial, mais especificamente, no começo dos anos 1970, o mundo passa por diversas transformações. Inicia-se a chamada Revolução Verde, no qual muda completamente a forma como a produção agrícola era realizada há séculos. Há o surgimento de maquinários, irrigação, produtos químicos oriundos da guerra são

utilizados com o intuito de controlar pragas, doenças e plantas daninhas (NUNES, 2007). O Brasil aproveitou o momento mundial, estimulou o crescimento de suas agroindústrias, principalmente no beneficiamento de café, cana-de-açúcar, soja e laranja, junto ao crescente aumento de produtividade das áreas agrícolas, elevando suas exportações (SOUZA, 2017).

Em meados dos anos 80 surge a agricultura de precisão (AP), juntamente com a Revolução Verde foi um grande passo para a produção agrícola mundial. A AP se utiliza de técnicas de produção agrícola, softwares para elaboração de projetos, programação de máquinas com técnicas já utilizadas em indústrias e principalmente a inserção do GPS (*Global Positioning System*) nos computadores de bordo do maquinário agrícola. No Brasil a agricultura de precisão tem sua chegada no início dos anos 2000, quando o governo cria projetos de incentivo à modernização do maquinário agrícola e as marcas trazem seus produtos com a tecnologia embarcada (BERNARDI et al., 2014).

A *smart agriculture* surge da evolução da agricultura de precisão, mantendo sua base e acrescentando novos equipamentos e tecnologias. Seu foco principal é facilitar a conectividade e obtenção de dados a fim de gerar informações úteis ao produtor rural (VILLAFUERTE et al., 2018). Diversas são as ferramentas utilizadas para obtenção dos dados para processamento, dentre elas estão drones, imagens de satélites, sensores, GPS e softwares de gestão (TORRES, 2017) agregando as novas tecnologias computacionais como *Big Data*, *Cloud Computing*, *Internet of Things* e *Machine Learning* (VILLAFUERTE et al., 2018).

Os autores BARMPOUNAKIS et al. (2015), afirmam que

“agricultura inteligente, agricultura de precisão, sistemas de informações de gerenciamento de fazendas (FMIS), todos são termos, que introduzem a noção de processos sofisticados utilizando informações avançadas, ferramentas de comunicação e sistemas nos processos agrícolas”.

A computação ubíqua expandiu as possibilidades de colaboração nos mais diversos segmentos da sociedade (FILIPPO et al., 2019). A facilidade da obtenção de diferentes objetos móveis, nos quais se conectam através de redes sem fio e os sistemas embarcados promoveram uma grande evolução na forma como a agricultura é conduzida (BERNARDI et al., 2014).

No contexto de *smart farming* todas as principais atividades de uma fazenda são monitoradas através de sensores espalhados pela propriedade e no maquinário utilizado no dia a dia. Os dados gerados são enviados para que possam ser processados e transformados em informações por sistemas gerenciais, geralmente de maneira automatizada (BARMPOUNAKIS et al., 2015). Com base nas informações obtidas, a

agricultura pode ser melhor gerenciada, promovendo maior eficiência e reduzindo custos e impactos ambientais (PIVOTO et al., 2016).

### **3. *Big Data* no agronegócio**

Durante a crescente demanda pela agricultura de precisão, facilitada pelo barateamento dos equipamentos e softwares embarcados, as fabricantes começaram a se preocupar com a confiabilidade dos dados, facilidade de utilização e instalação nas máquinas agrícolas, não levando em consideração uma possível integração em base de dados. Dentro dos maquinários havia uma grande quantidade de terminais e equipamentos de diferentes empresas para a obtenção de diferentes informações, o que não era muito produtivo (BERNARDI et al., 2014).

Reconhecida esta demanda, forças tarefa americanas e europeias se uniram para a criação de uma ISO para padronização de comunicação e transferência de informações entre máquinas agrícolas, a ISO 11783, conhecida também como ISOBUS (GEO AGRI, 2019). Essa ISO segue um padrão de comunicação digital criado nos anos 80, *Controller Area Network* (CAN), que inicialmente era utilizado no setor automotivo e logo migrou para outras áreas, desde construção civil até equipamentos militares e de exploração aeroespacial (SOUZA et al., 2007). Após a padronização, um implemento pode se conectar a um trator através de *Plug and Play* e já terá suas informações visualizadas em um terminal universal facilitando a coleta das informações (GEO AGRI, 2019).

A infraestrutura das propriedades está sendo alterada, cada vez mais componentes eletrônicos presentes no dia a dia do produtor rural. Redes sem fio conectadas a internet com componentes ubíquos tão rotineiros que acabam se tornando invisíveis e aproximam cada vez os atores do contexto agrícola (BERNARDI et al., 2014). Além dos dispositivos citados, sequenciamentos de genomas e redes de sensores (SOUZA et al., 2017) espalhados a nível do solo, nível aéreo e nível orbital (GIRALDELLI, 2020) geram uma quantidade imensa de dados de grande variedade e com coleta cada vez mais rápida (LEITE et al., 2014).

A geração de dados tem aumentado exponencialmente na sociedade como um todo, no agronegócio não é diferente, estima-se que a cada 10 anos o volume de dados gerados tem aumentado por fator de 100, o que implica em uma subutilização destes. Unificando esses dados em um mesmo local, *Big Data*, é possível realizarmos estudos utilizando técnicas computacionais, matemáticas, estatísticas entre outras, buscando encontrar relações nunca antes imaginadas, já que somente são possíveis de realizar devido a evolução tecnológica dos equipamentos (FISCHER et al., 2020).

O *IoT (Internet of Things)* proporciona ao agronegócio a obtenção dos dados em tempo real, seu ecossistema é composto por sensores, *hardware*, *software* e a capacidade de conexão a redes sem fio, que proporcionaram o envio dos dados coletados. Um exemplo de utilização com envio constante dos dados é a utilização em um silo de armazenamento de grãos no qual instalado um ecossistema *IoT* no topo, com câmeras e sensores de gases como o CO<sub>2</sub>, é possível obter a quantidade armazenada e a qualidade dos grãos (OLIVEIRA, 2020).

Montone e Russo (2018) estimam que uma fazenda possa gerar em torno de 20gb de dados por safra, e durante este período um produtor tenha que tomar inúmeras decisões, desde a escolha das sementes, época de aplicações de insumos, quais insumos entre outras. Sonka (2020) diz que variedade e análise dos dados são dois fatores comumente ignorados em um montante de *Big Data*, sem uma análise bem feita são apenas números sem qualquer agregação de valor.

Os dados agrícolas possuem os mais diversos modelos e formatos, a gama de objetos que geram informações é muito grande e desigual (SARAIVA et al., 2020), os dados podem representar desde a qualidade química e física do solo, dados sobre o clima, cotações de *commodities*, aplicações de insumos agrícolas, níveis de infestação de pragas e doenças, cultivares, produtividade entre outros (SOUZA, 2018). Devido a essa diversidade, qualidade despadronizada e o tempo acelerado no qual os dados são gerados, modelos tradicionais estatísticos não são capazes de processar todo o montante, sendo uma atividade multidisciplinar envolvendo profissionais de áreas matemáticas, tecnológicas e no caso em questão, profissionais com conhecimento de negócio agrícola como o engenheiro agrônomo, zootecnista, engenheiro florestal entre outros (SARAIVA et al., 2020).

#### **4. *Open Source***

Entre as décadas de 50 e 60, a rede *Advanced Research Projects Agency Network* (ARPANET) iniciava a criação de uma rede de comunicação e protocolos de Internet, era comum que os desenvolvedores compartilhassem suas pesquisas para que fossem revisadas e melhoradas. Com o surgimento da Internet moderna como conhecemos nos dias atuais, essa cultura de abertura e compartilhamento de código já era a anos utilizada por pesquisadores e desenvolvedores que promoveram ainda mais sua adoção (RED HAT, 20-?).

Esse movimento de compartilhamento de código que permite a possibilidade de visualização, modificação e distribuição é conhecido como *Open Source*. Os softwares

*Open Source* geralmente têm uma vida de utilização prolongada, já que é mantido de forma descentralizada e permite que comunidades de desenvolvedores possam trabalhar de forma independente sempre melhorando-o (RED HAT, 20-?).

#### 4.1. *Open Data*

O conceito de *Open Data* ou dados abertos segue o conceito de *Open Source*, os dados abertos são disponibilizados de maneira que sua obtenção seja facilitada, preferencialmente pela Internet. Seus termos devem permitir sua reutilização e redistribuição mesmo que associados a outras bases de dados. Caso o *dataset* disponibilizado, em seus termos de uso, informe que poderá ser usado somente em determinado segmento, como educativo, este não é mais considerado *Open Data* (OPEN DATA HANDBOOK, 20-?).

Empresas privadas e públicas têm utilizado a abertura de dados e informações para construir suas estratégias, tornando o acesso à informação como alavanca de crescimento e desenvolvimento no século XXI (SANTARÉM SEGUNDO, 2015).

Com a popularização do conceito de *Open Data*, governos de diversos países do mundo começaram a disponibilizar dados em plataformas próprias, o chamado *Open Government*. O governo dos EUA (2019) define os dados como

“um recurso nacional valioso e um ativo estratégico para o governo dos EUA, seus parceiros e o público. Gerenciar esses dados como um ativo e torná-los disponíveis, detectáveis e utilizáveis - em uma palavra, abertos - não apenas fortalece nossa democracia e promove a eficiência e eficácia do governo, mas também tem o potencial de criar oportunidades econômicas e melhorar a qualidade dos cidadãos de vida”.

O governo brasileiro (BRASIL, 20-?) se baseia na definição da *Open Knowledge Internacional* sobre dados abertos que diz

“dados são abertos quando qualquer pessoa pode livremente acessá-los, utilizá-los, modificá-los e compartilhá-los para qualquer finalidade, estando sujeito a, no máximo, a exigências que visem preservar sua proveniência e sua abertura”.

O *Open Data Barometer* é uma organização internacional que visa a classificação dos países do mundo em relação a publicação de dados abertos, em sua 1ª edição realizada no ano de 2013, o Brasil ocupava a 28ª posição, já na 4ª e última publicação do ranking, passou a ocupar a 18ª posição (OPEN DATA BAROMETER, 2016).

No ano de 2007, no estado da Califórnia nos EUA, foi realizada uma conferência que reuniu um grupo de 30 advogados de diversos países. Nela foram definidos 8 princípios para abertura de dados governamentais. São eles: os dados devem ser completos, não tendo validade ou questões de privacidade. Devem ser primários, ou seja, não podem ser transformados ou receberem adição de outros *datasets*. Devem ser atuais, no momento

que são coletados já são disponibilizados. Devem ser acessíveis, facilmente encontrados na Internet. Devem ser estruturados para que possam ser processados por máquinas. Acessados por qualquer cidadão, sem discriminação, não necessitando de cadastros. Devem ser disponibilizados em formatos abertos. E devem possuir licença livre, não podendo ser patenteado ou ser caracterizado como segredo industrial (OPEN GOVERNMENT DATA, 2007).

O potencial de geração de informação através de *Open Data* e *Open Data Government* pode ser aumentado através de técnicas como *Linked Open Data* (LOD), *Resource Description Framework* (RDF) (KLEIN, 2015) e *Uniform Resource Identifiers* (URIs) (MARCONDES, 2012), criando ligações semânticas entre os dados (KLEIN, 2015).

#### **4.2. Web Semântica e *Linked Open Data***

Desde a popularização da Internet para o mundo, no início dos anos 1990, um entrave é debatido entre especialistas, como estruturar os dados na *Web*. Essa discussão surge porque a Internet foi iniciada como um grande depósito de dados e informações nos quais havia grande dificuldade em encontrá-los (CONEGLIAN & SANTARÉM, 2016).

Em 2001, Tim Berners-Lee, James Handler e Ora Lassila publicaram um artigo no qual propõem tecnologias para a criação da base da *Web* semântica, visando facilitar a organização e estruturação da *Web* para que computadores consigam encontrar e compreender melhor os dados (CONEGLIAN & SANTARÉM, 2016).

Na *Web* “tradicional”, softwares buscam informações através da forma sintática dos dados, ou seja, são buscados dados idênticos aos informados para a procura. A *Web* semântica propõe que ao invés de serem buscados os dados idênticos ao informado, softwares sejam capazes de localizar diferentes tipos de dados com amplas possibilidades de interações (MARCONDES, 2012), ela torna capaz de transformar o conteúdo da *Web* sintática em conteúdos com significado (SILVEIRA et al., 2020).

O *Linked Open Data* (LOD) visa melhorar a *Web* semântica, explorando-a, possuindo poucos dados iniciais, é possível se conectar a diferentes fontes de dados ao redor do mundo, formando uma rede única de informações “linkadas” pelos nós que o LOD proporciona (SANTARÉM SEGUNDO, 2015). Segundo Heath (2012, apud SANTARÉM SEGUNDO, SOUZA & CONEGLIAN, 2015), *Linked Data* “é um conjunto de melhores práticas para publicação e conexão de dados estruturados na *Web*, permitindo estabelecer links entre itens de diferentes fontes de dados para formar um único espaço de dados global”.

Em 2006, Berners-Lee publicou quatro princípios para a construção do LOD:

- a. Usar URIs para nomear coisas;
- b. Usar HTTP URIs para facilitar que pessoas encontrem esses nomes;
- c. Quando alguém pesquisar por uma URI fornece um conteúdo relevante usando padrão RDF;
- d. Incluir mais *links* relacionados para facilitar a descoberta de novos itens.

O *Uniform Resource Identifier* (URI) é definido por Berners-Lee (2005) como “uma sequência compacta de caracteres que identifica um recurso abstrato ou físico”. Os autores Santarém Segundo, Souza e Coneglian (2015) descrevem a URI como um identificador único, representando qualquer tipo de dado na *Web*, algo muito importante, já que se o recurso for confundido com outros, pode produzir informações incorretas.

Outra técnica utilizada para o LOD é o *Resource Description Framework* (RDF), que é um modelo de estrutura, utilizando na grande maioria dos casos a *Extensible Markup Language* (XML) como sintaxe para descrever os objetos na *Web*, possuindo três elementos básicos, recursos, propriedades e declarações (FERREIRA & SANTOS, 2013). Descrevendo cada elemento, temos o recurso como a identificação do objeto de dado, a URI; as propriedades podem ser descritas como as características que distinguem os recursos, permitindo sua correta identificação e não permitindo a criação de informações incorretas e; a declaração que é objeto como um todo, com recurso e propriedades (SANTARÉM SEGUNDO, SOUZA & CONEGLIAN, 2015). O principal objetivo do RDF é permitir a troca de metadados entre sistemas computacionais preservando sua originalidade (FERREIRA & SANTOS, 2013).

## 5. Visualização de Dados

A visualização de dados através de gráficos existe a séculos, no século XVII iniciaram-se as criações de mapas e diagramas culminando no início do século XVIII no primeiro gráfico de pizza. Décadas depois, criada por Charles Minard, a representação da invasão da Rússia por Napoleão foi uma das representações gráficas mais conhecidas da história. Minard conseguiu representar o tamanho do exército de Napoleão e o caminho de retirada deste exército de Moscou, relacionando com a temperatura e a escala de tempo (SAS, 2019).

A espécie humana é extremamente visual, desde os desenhos em paredes nos primórdios das civilizações, a arte em seus segmentos como quadros e esculturas, até a TV e cinema nos dias atuais. Rapidamente nossos olhos conseguem diferenciar as cores e diferentes formas geométricas (TABLEAU, 2020).

A visualização de dados tem a capacidade de atrair melhor a atenção das pessoas em relação ao que está sendo representado (TABLEAU, 2020). A capacidade de manipulação da representação, faz com que o entendimento e o encontro de diferentes padrões sejam facilmente observados pelos usuários. Essa facilidade ocorre porque relacionamos os dados com algo concreto, como barras ou linhas (FREITAS et al., 2001).

## **6. Proposta de Desenvolvimento do Trabalho**

### **6.1. Objetivos**

Esta pesquisa de iniciação científica tem como objetivo principal estudar e demonstrar como aplicar o Data Science dentro do contexto agrícola, utilizando tecnologias *open source* para análise e visualização de dados obtidos durante as atividades agrícolas.

### **6.2. Tecnologias e Recursos *Open Source***

#### **6.2.1. Python**

Python é uma linguagem de programação de alto nível, robusta, orientada à objetos, interpretada via *bytecode*, multiplataforma e *Open Source*. Não possui um objetivo definido, podendo ser utilizada para criação de aplicações *web*, *desktop*, aplicativos móveis, protocolos de rede, jogos e para ciência de dados no qual foi onde cresceu sua popularidade (PYTHON BRASIL, [entre 2017 e 2020]).

Em relatório divulgado pela IEEE *Spectrum* em julho de 2020, Python é a linguagem de programação mais popular do mundo até data de divulgação do estudo (CIO, 2020). O Python se tornou popular em projetos que têm a necessidade de manipulação de grandes massas de dados, ela tem a capacidade de trabalhar com grandes números e realizar cálculos complexos com os mesmos. Tem forte atuação em aplicações de *Machine Learning*, *Deep Learning*, inteligência artificial, computação 3d, *Data Science*, *Big Data* entre outros (SILVA, 2018). Entre as bibliotecas mais conhecidas para estes usos estão: NumPy, Pandas e Matplotlib, (SANTANA, 2019).

#### **6.2.2. NumPy**

NumPy é uma importante biblioteca para o ecossistema *Data Science* utilizando a linguagem Python. É capaz de simplificar a realização de cálculos numéricos em *arrays* multidimensionais (SANTIAGO JUNIOR, 2018).

Essa biblioteca ganhou popularidade por trazer todo o poder computacional de linguagens como C e Fortran para o Python, tornando-se simples e funcional, já que a linguagem Python tem uma curva de aprendizado relativamente menor (NUMPY, 2019-2020).

Devido a sua facilidade de escrita e velocidade no processamento dos cálculos, NumPy é muito utilizada em *Machine Learning*, o treinamento dos algoritmos é abastecido com dados armazenados em *arrays* multidimensionais, o processo de aprendizado de máquina é agilizado com sua utilização. Outra tarefa muito importante que é otimizada utilizando NumPy é a manipulação de imagens e computação gráfica, as imagens também são representadas com *arrays* (SANTIAGO JUNIOR, 2018). Outros exemplos de utilização citados pela NumPy: computação quântica, estatística computacional, processos astronômicos, psicologia cognitiva, bioinformática, análises multivariáveis, processamento geográfico entre outros (2019-2020).

### **6.2.3. Pandas**

Pandas é outra importante biblioteca para *Data Science*, tem papel fundamental na análise e manipulação dos dados. Tem a capacidade de trabalhar com qualquer tipo de arquivo, exemplos: csv, excel, txt, etc (SANTANA, 2019).

De acordo com a Pandas Org. seu objetivo é ser facilmente utilizada por qualquer pessoa, requerendo apenas conhecimentos básicos na linguagem Python (PANDAS, 2020). Essa biblioteca possui três funcionalidades básicas para análise de dados que são: manipulação, leitura e visualização dos dados (FIGUEIREDO, 2018).

Os dados podem ser estruturados de duas maneiras, *Dataframe* ou *Series* (SANTANA, 2019). O *Dataframe* possui estrutura de um *array* multidimensional, onde as colunas representam os campos e as linhas os registros, como em uma tabela de banco de dados relacional (SANTANA, 2019). Já a estrutura *Series* é mais simples, possuindo apenas uma chave e uma coluna (SANTANA, 2019).

### **6.2.4. Matplotlib**

Pela definição do Matplotlib Team (2012), Matplotlib “é uma biblioteca abrangente para a criação de visualizações estáticas, animadas e interativas em Python”.

Essa biblioteca é a mais famosa para visualização das informações obtidas através da análise dados, outras bibliotecas como Bokeh e Seaborn, a utilizam como base (VASCONCELLOS, 2017). Basicamente ela trabalha plotando gráficos 2d, com eixo x e y. Utiliza também a biblioteca NumPy para melhorar seu desempenho na criação de

gráficos, gráficos de barras, de dispersão, histogramas entre outros (MATPLOTLIB, 2012).

### **6.2.5. Jupyter**

O Jupyter é um ambiente de desenvolvimento *Open Source* e interativo, pode ser utilizado via *Web* ou instalado localmente. Ele é altamente produtivo, permitindo que cada trecho de código possa ser executado independentemente e ainda ser descrito, já que aceita a inserção de texto em qualquer local do documento (VIANA, 2017).

A possibilidade de criação de narrativas computacionais reproduzíveis fez com que esta ferramenta se tornasse muito popular entre cientistas de dados, já que intercalam textos explicativos, gráficos e tabelas dinâmicas (SCARDINE, 2018). Além de ser extensível e modular permitindo que qualquer pessoa possa criar *plugins* e disponibilizá-los para outros usuários (JUPYTER, 2020).

### **6.3. Testes e Resultados**

Utilizando os recursos supra citados, foi possível verificar que com conhecimentos básicos das tecnologias é possível manipular os dados e criar visualizações que os transformam em informações relevantes para tomadas de decisões.

As ferramentas possuem vasta documentação sobre suas funcionalidades, facilitando a busca de informação para sua correta utilização, com elas foi possível construir visualizações sobre as situações socioeconômicas dos produtores rurais brasileiros, além da construção de mapas de produtividade com dados abertos obtidos a partir de sites governamentais.

### **6.4. Trabalhos Futuros**

Para trabalhos futuros podem ser sugeridos diversos temas abrangendo *Data Science*, dentre eles o estudo de algoritmos de *Machine Learning* aplicados em dados abertos.

## **7. Conclusões**

O montante de dados gerados em todos os setores da economia mundial é muito alto, o cientista de dados inglês, Clive Humby, diz que “Data is the new oil” – dados são o novo petróleo. E quem conseguir utilizá-los de forma correta, fazendo bom uso para aproveitamento total de seu potencial, sairá na frente dos concorrentes.

Dessa forma, essa pesquisa propôs estudar, analisar e estimular novas pesquisas de utilização do Data Science na análise e visualização de dados gerados pela agricultura

com ferramentas *open source*, para que a atividade se torne cada vez produtiva e sustentável.

## 8. REFERENCIAS BIBLIOGRÁFICAS

BARMPOUNAKIS, S. et al. **Management and control applications in Agriculture domain via a Future Internet Business-to-Business platform**. Research Gate, ano 2015. Disponível em: < <https://cutt.ly/Rg5Ruqq> >. Acesso em: 18 JUN 2020.

BERNARDI, A. C. D. C. et al. **Agricultura de precisão: resultados de um novo olhar**. Empresa Brasileira de Pesquisa Agropecuária (Embrapa), ano 2014. Disponível em: < <https://cutt.ly/Eg5Riof> >. Acesso em: 16 JUN 2020.

BERNERS-LEE, T. **Linked data principles**. Ano 2006. Disponível em: < <https://cutt.ly/bg5Roj3> > Acesso em: 21 AGO 2020.

BERNERS-LEE, T. **Uniform Resource Identifier (URI): Generic Syntax**. Ano 2005. Disponível em: < <https://cutt.ly/Ug5RpnS> >. Acesso em: 21 AGO 2020.

BRASIL. **Portal Brasileiro de Dados Abertos**. [20-?]. Disponível em: < <https://cutt.ly/cg5Rp39> >. Acesso em 09 AGO 2020.

CIO. **Python domina ranking de linguagens de programação de 2020**. Ano 2020. Disponível em: < <https://cutt.ly/eg5RaAe> >. Acesso em 02 SET 2020.

CONGLIAN, C. S.; SANTARÉM, J. E. **Europeana no Linked Open Data: Conceitos de Web Semântica na dimensão aplicada das Humanidades Digitais**. Florianópolis, Encontros Bibli: revista eletrônica de biblioteconomia e ciência da informação, vol. 22, núm. 48, ano 2017. Disponível em: < <https://cutt.ly/Vg5RszO> >. Acesso em: 10 AGO 2020.

FAO. **E-Agriculture in action: big data for agriculture**. 2019. Disponível em: < <https://cutt.ly/Vg5RdLx> >. Acesso em: 20 nov. 2019.

FERREIRA, J. A.; SANTOS, P. L. V. A. da C. **O modelo de dados Resource Description Framework (RDF) e o seu papel na descrição de recursos**. João Pessoa, Informações & Sociedade, vol. 23, núm. 2, ano 2013. Disponível em: < <https://cutt.ly/wg5Rg1f> >. Acesso em: 10 AGO 2020.

FILHO, P. F.; PAULA, S. **A Agroindústria. Banco Nacional de Desenvolvimento Econômico e Social (BNDES)**, ano 2001. Disponível em: < <https://cutt.ly/Sg5RhC0> >. Acesso em: 09 JUL 2020.

FILIPPO, D. et al. **Sistemas Colaborativos**. Uniriotec, ano 2019. Disponível em: < <https://cutt.ly/9g5RjUO> >. Acesso em: 16 JUN 2020.

FISCHER, H.; SALVI, J. V.; GARCIA, L. H. T. **Transmissão de dados, computação nas nuvens e Big Data**. In: QUEIROZ, D. M.; VALENTE, D. S. M.; PINTO, F. A. C.; BORÉM, A. *Agricultura Digital*. Viçosa: Universidade Federal de Viçosa (UFV), 2020.

FREITAS, C. M. D. S.; CHUBACHI, O. M.; LUZZARDI, P. R. G.; CAVA, R. A. **Introdução à visualização de informações**. Porto Alegre. Revista de informática teórica e aplicada, vol. 8, núm. 2, ano 2001. Disponível em: < <https://cutt.ly/pg5RkRy> >. Acesso em 12 SET 2020.

GEO AGRI. **ISOBUS - Comunicação entre tratores e implementos agrícolas**. Geo Agri Tecnologia Agrícola, ano 2019. Disponível em: < <https://cutt.ly/Fg5RlsZ> >. Acesso em: 15 JUL 2020.

GIRALDELLI, A. L. **Sensoriamento remoto na agricultura: 7 coisas que você deveria saber**. Lavoura 10, ano 2020. Disponível em: < <https://cutt.ly/2g5R11Q> >. Acesso em: 20 JUL 2020.

JUPYTER. JUPYTER. Ano 2020. Disponível em: < <https://cutt.ly/kg5RzuN> >. Acesso em: 04 SET 2020.

KLEIN, V. B. **Uma proposta de modelo conceitual para uso de Big Data e Open Data para Smart Cities**. 2015. 167 p. Dissertação (Mestrado em Engenharia e Gestão do Conhecimento) - Universidade Federal de Santa Catarina, Florianópolis, 2015. Disponível em: < <https://cutt.ly/Zg5RzFV> >. Acesso em: 11 AGO 2020.

LEITE, M. A. A. et al. Tecnologias emergentes - futuro e evolução tecnológica das AgrotIC. In: MASSRUHÁ, S. M. F. S. et al. **Tecnologias da informação e comunicação e suas relações com a agricultura**. Brasília: EMBRAPA, ano 2014, cap. 17. Disponível em: < <https://cutt.ly/sg5RxEa> >. Acesso em: 20 JUL 2020.

MAPA. **Principais conceitos e expressões utilizados na agricultura de precisão - AP**. Ministério da Agricultura, Pecuária e Abastecimento, Brasília, ano 2010. Disponível em: < <https://cutt.ly/yg5Rca6> >. Acesso em: 13 JUN 2020.

MARCONDES, C. H. **“Linked Data” - Dados interligados - e interoperabilidade entre arquivos, bibliotecas e museus na Web**. Florianópolis, Encontros Bibli: revista eletrônica de biblioteconomia e ciência da informação, vol. 17, núm. 34, ano 2012. Disponível em: < <https://cutt.ly/hg5RvQa> >. Acesso em: 10 AGO 2020.

MATPLOTLIB. **Matplotlib: Visualization with Python**. Site, ano 2012. Disponível em: < <https://matplotlib.org/index.html> >. Acesso em: 03 SET 2020.

MONTONE, V., RUSSO, D. **O “bem-casado” da ciência de dados e da agricultura**. Blog Climate FieldView, ano 2018. Disponível em: < <https://cutt.ly/2g5Rv4M> >. Acesso em: 20 JUL 2020.

NUMPY. **Ecosystem**. Site, ano 2019-2020. Disponível em: < <https://cutt.ly/Bg5RbKl> >. Acesso em: 03 SET 2020.

NUNES, S. P. **O desenvolvimento da agricultura brasileira e mundial e a ideia de desenvolvimento rural**. Departamento de Estudos Sócio-Econômicos Rurais (DESER), nº 157, ano 2007. Disponível em: < <https://cutt.ly/dg5Rnst> >. Acesso em: 09 JUL 2020.

OLIVEIRA, H. S. Internet das coisas para agricultura. In: QUEIROZ, D. M.; VALENTE, D. S. M.; PINTO, F. A. C.; BORÉM, A. **Agricultura Digital**. Viçosa: Universidade Federal de Viçosa (UFV), 2020.

ONU. **World Resources Report: Creating a sustainable food future**. 2019. Disponível em: <https://cutt.ly/Xg5Rn34>. Acesso em: 20 nov. 2019.

OPEN DATA BAROMETER. **A global measure of how governments are publishing and using open data for accountability, innovation and social impact**. 2016. Disponível em: < <https://cutt.ly/Yg5Rmc2> >. Acesso em 09 AGO 2020.

OPEN DATA HANDBOOK. **O que são dados abertos?** [20-?]. Disponível em: < <https://cutt.ly/Ug5Rm1Y> >. Acesso em: 07 AGO 2020.

OPEN GOVERNMENT DATA. **Open Government Data Principles**. Califórnia, 2007. Disponível em: < <https://cutt.ly/Kg5RQml> >. Acesso em: 09 AGO 2020.

PANDAS. **Pandas**. Site, ano 2020. Disponível em: < <https://cutt.ly/Pg5RQ2g> >. Acesso em 03 SET 2020.

PEREIRA, W. **Há espaço para uma relação harmoniosa entre a produção de biocombustíveis e a segurança alimentar?** Disponível em: <https://cutt.ly/6g5RWEk>. Acesso em: 20 nov. 2019.

PIVOTO, D. et al. **Smart Farming pode trazer ganhos ao agronegócio?** Biblioteca Digital FGV, ano 2016. Disponível em: < <https://cutt.ly/mg5REP8> >. Acesso em: 18 JUN 2020.

PYTHON BRASIL. **Python para quem está começando**. Blog Python Brasil, ano [entre 2017 e 2020]. Disponível em: < <https://cutt.ly/bg5RRUN> >. Acesso em: 02 SET 2020.

RED HAT. **O que é Open Source?** [20-?]. Disponível em: < <https://cutt.ly/7g5RR41> >. Acesso em: 07 AGO 2020.

SANTANA, F. **Análise de dados com Python usando Pandas**. Blog Minerando Dados, ano 2019. Disponível em: < <https://cutt.ly/Dg5RYtE> >. Acesso em: 03 SET 2020.

SANTANA, F. **Confira as melhores bibliotecas da linguagem Python**. Blog Minerando Dados, ano 2019. Disponível em: < <https://cutt.ly/Gg5RYIA> >. Acesso em: 02 SET 2020.

SANTARÉM SEGUNDO, J. E. **Web semântica, dados ligados e dados abertos: uma visão dos desafios do brasil frente às iniciativas internacionais**. XVI ENANCIB - Encontro Nacional de Pesquisa em Ciência da Informação. João Pessoa, 2015. Disponível em: < <https://cutt.ly/GhtVuko> >. Acesso em: 10 AGO 2020.

SANTARÉM SEGUNDO, J. E.; SOUZA, J. O.; CONEGLIAN, C. S. **Web semântica: Introdução a recursos de visualização de dados em formato gráfico**. XVI ENANCIB - Encontro Nacional de Pesquisa em Ciência da Informação. João Pessoa, 2015. Disponível em: < <https://cutt.ly/Kg5RUIA> >. Acesso em: 10 AGO 2020.

SANTIAGO JUNIOR, L. **Entendendo a biblioteca NumPy**. Blog Medium, ano 2018. Disponível em: < <https://cutt.ly/ug5RU3l> >. Acesso em: 03 SET 2020.

SARAIVA, A. M. et al. Dados digitais: ciclo, padronização, qualidade, compartilhamento e segurança. In: QUEIROZ, D. M.; VALENTE, D. S. M.; PINTO, F. A. C.; BORÉM, A. **Agricultura Digital**. Viçosa: Universidade Federal de Viçosa (UFV), 2020.

SAS. **Visualização de dados – O que é e qual sua importância?** Ano 2019. Disponível em: < <https://cutt.ly/eg5ROqk> >. Acesso em: 12 SET 2020.

SCARDINE, P. **JupyterLab: a nova geração de Jupyter Notebooks**. Blog Medium, ano 2018. Disponível em: < <https://cutt.ly/5g5RPwy> >. Acesso em: 04 SET 2020.

SILVA, D. M. **Python: História e essência**. Revista Programar, ed. 59, p. 96 – 99, ano 2018. Disponível em: < <https://cutt.ly/dg5RPb3> >. Acesso em: 02 SET 2020.

SILVEIRA, L. da et al. **Utilização De Linked Open Data Em Bibliotecas: O Caso Da Bibliothèque Nationale De France**. Florianópolis, Revista ACB: Biblioteconomia em Santa Catarina, vol. 25, núm. 1, ano 2020. Disponível em: < <https://cutt.ly/ug5RPJv> >. Acesso em: 10 AGO 2020.

SONKA, S. Sem análise, dados não geram valor para o agro, afirma especialista em Big Data. In: COLUSSI, J. **Revista Globo Rural**, ed. 414, ano 2020. Disponível em: < <https://cutt.ly/0g5RAGi> >. Acesso em: 20 JUL 2020.

SOUZA, G. L. R. **História do Agronegócio no Brasil**. Folha Acadêmica do Centro de Ensino Superior de São Gotardo, nº XIII, ano 2017. Disponível em: < <https://cutt.ly/vg5RSwf> >. Acesso em: 09 JUL 2020.

SOUZA, K. X. S. et al. **Tecnologia da informação na agropecuária - estado da arte, tendências futuras e proposta de atuação**. Campinas: EMBRAPA, ano 2017. Disponível em: < <https://cutt.ly/Zg5RSbM> >. Acesso em: 20 JUL 2020.

SOUZA, R. V.; GODOY, E. P.; PORTO, A. J. V.; INAMASU, R. Y. **Redes embarcadas em máquinas e implementos agrícolas: o Protocolo CAN (Controller Area Network) e a ISO 11783 (ISOBUS)**. São Carlos: EMBRAPA, ano 2007. Disponível em: < <https://cutt.ly/Wg5RDno> >. Acesso em: 27 JUL 2020.

TABLEAU. **Guia prático da visualização de dados: definição, exemplos e recursos de aprendizado**. Ano 2020. Disponível em: < <https://cutt.ly/Mg5RDNt> >. Acesso em: 12 SET 2020.

TORRES, A. B. B. **Fusão de dados multinível para sistemas de internet das coisas em agricultura inteligente**. Fortaleza, ano 2017. Disponível em: < <https://cutt.ly/Yg5RFdT> >. Acesso em: 13 JUN 2020.

U.S. GENERAL SERVICES ADMINISTRATION. **Project Open Data, Open Data Policy Managing Information as an Asset**. 2019. Disponível em: < <https://cutt.ly/Hg5RF2c> >. Acesso em: 08 AGO 2020.

VIANA, S. **Por que usar Jupyter Notebook?** Blog Medium, ano 2017. Disponível em: < <https://cutt.ly/lg5RGIE> >. Acesso em: 04 SET 2020.

VILLAFUERTE, A. et al. **Agricultura 4.0 - Estudo De Inovação Disruptiva No Agronegócio Brasileiro**. 9th International Symposium on Technological Innovation, Aracaju-SE, p. 150-162, ano 2018. Disponível em: < <https://cutt.ly/zg5RGHg> >. Acesso em: 15 JUL 2020.