|  |  |
| --- | --- |
| PIBIFSP | PROJETO DE PESQUISA |

|  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| TÍTULO DO PROJETO:  A INTELIGÊNCIA ARTIFICIAL APLICADA NA INTERAÇÃO HUMANO COMPUTADOR: UM ESTUDO APLICADO AO DESENVOLVIMENTO DE UM CHATBOT PARA O IFSP CAMPUS VOTUPORANGA | | | | | | | | | | | | |
| Área do Conhecimento (Tabela do CNPq): | 6 | **.** | 0 | 7 | **.** | 0 | 2 | **.** | 0 | 3 | **-** | 6 |

**1 RESUMO**

Esse projeto de pesquisa visa o estudo das ferramentas, tecnologias, teorias e metodologias para o desenvolvimento e a aplicação de um *Chatbot*, o qual, terá uma aplicação para o atendimento de usuários que procuram entrar em contato com o IFSP Campus Votuporanga. Os *Chatbots* são programas utilizados para conversações entre sistemas e humanos, e como ferramentas produtivas que facilitam e agilizam as atividades entre os indivíduos. Assim, a partir dos conceitos de *Machine Learning* e processamento de linguagem natural, o sistema terá como objetivo a automatização na mediação do contato entre o campus e os docentes, discentes ou a comunidade externa através das redes sociais. O algoritmo permitirá a implementação de novas respostas a base de informações, que podem ser consultadas pelo usuário posteriormente, o sistema ainda contará com um fluxo de simulação.

**2 FUNDAMENTAÇÃO TEÓRICA**

Nessa seção pretende-se apresentar a fundamentação teórica para o desenvolvimento dessa pesquisa. A metodologia utilizada é de cunho teórico, prático, exploratório. Almeja-se explorar as possibilidades que envolvem nosso problema científico para assim reunir conteúdos teóricos capazes de identificar os rumos em relação às metodologias, técnicas e práticas a partir do *Machine Learning* e do processamento de linguagem natural.

**2.1 Contextualização**

Com o crescente avanço da tecnologia trazendo inovações em aparelhos eletrônicos e suas novas formas de interação com os seres humanos, o mercado consumidor é estimulado ainda mais e, somando-se ao aumento do poder de consumo por parte da população em geral, criou-se um ecossistema de constante produção de bens de consumo que não se viam a algumas décadas atrás, e suas principal diferença é o emprego de técnicas mais sofisticadas junto de hardwares e softwares. Como consequência surgiram novas áreas de estudo, como Inteligência Artificial, Biofísica Computacional, entre outros.

Nesse ambiente de constante reformulação criou-se uma demanda por profissionais capazes que possam se adequar ao meio, que disponibilizem seus conhecimentos e os apliquem, sendo o principal enfoque de empresas de grande porte. Mas também, é considerado extremamente benéfico aos usuários que compreendam ao menos conhecimentos básicos, principalmente as crianças que serão a próxima geração de pessoas que irão utilizar estas tecnologias no cotidiano e estarão envoltos por fluxos de informações.

Sendo a base destes conhecimentos e o primeiro contato com este “mundo” - a lógica de programação, que pode ser aplicada nos mais diversos ramos da ciência da computação e que a maioria dos cursos correlacionados trazem na sua grade curricular a disciplina de algoritmos, normalmente aplicada no primeiro ano. Entretanto, os alunos que pretendem iniciar na área muitas vezes demonstram dificuldades ao decorrer das aulas, em que não conseguem desenvolver o pensamento computacional, e por vezes se sentem desmotivados, culminando na evasão da disciplina ou até mesmo, na desistência do curso.

Pensando-se nesse problema, surge como necessidade a busca por novos métodos, tecnologias e ferramentas que possam auxiliar no ensino-aprendizagem do pensamento lógico, fazendo com que este primeiro contato ocorra de uma forma mais instigante e desafiadora, e que forneça um escopo maior de ângulos de visão para o aluno e perceba as mais diversas aplicações que o seu conhecimento adquirido pode proporcionar.

**Ensino tradicional de Algoritmos e Programação**

Atualmente, o ensino da lógica de programação e algoritmos no Brasil acontece principalmente em cursos superiores na área de Computação e informática, ou em cursos profissionalizantes.

De acordo com as Diretrizes Curriculares Nacionais para os cursos de graduação em Computação, lógica e algoritmos fazem parte dos conteúdos curriculares da formação tecnológica e básica para todos os cursos de bacharelado e de licenciatura (Ministério da Educação, 2012).

Geralmente esses conteúdos estão inseridos nas disciplinas introdutórias identificadas como: Lógica de programação, Algoritmos, Linguagens de programação, entre outros. E tem como objetivo introduzir ao aluno conceitos básicos de programação, como pensamento crítico, abstração de problemas, as sequências de passos para resolvê-los, verificação dos resultados, entre outros.

De modo geral, a ementa destas disciplinas iniciais contempla assuntos como a resolução de problemas utilizando algoritmos e raciocínio lógico, tipos de dados, variáveis e constantes, expressões, operadores de atribuição, matemáticos e lógicos, estruturas de controle, condicionais, estruturas de repetição; e estruturas básicas de dados, por exemplo, vetores, matrizes e funções.

Por estas disciplinas terem uma grande carga de conceitos abstratos, não sendo considerados de simples compreensão, os alunos iniciantes sentem dificuldades de assimilar e pôr em prática estes conceitos, e como consequência, serem um dos principais motivos pela alta taxa de evasão e reprovação nos cursos da área de Computação (Silva, Melo, & Tedesco, 2016).

Deve-se levar em conta também que o método de ensino apresentado na disciplina é centrado na transmissão de informação do professor, que é quem detém o aprendizado, para o aluno, através de conteúdos com embasamentos teóricos, exemplos e propostas de exercícios. Porém, esses métodos tradicionais na maioria dos casos não se adequam às necessidades dos alunos, visto que os professores por restrições temporais não conseguem dar o feedback e supervisão adequados e necessários para cada aluno, não conseguindo assim identificar e explorar facilidades e dificuldades que cada indivíduo possa vir a possuir no processo de aprendizagem da disciplina.

“Neste sentido, o ensino tradicional evidencia duas situações: estudantes habituados a serem indivíduos passivos dentro do ambiente escolar e estudantes/professores limitados pelo tempo” (AMARAL,2017). Reconhecendo que os métodos implantados no ensino de algoritmos possuem brechas e podem ser melhorados, começamos a identificar alternativas tecnológicas que possam aprimorar este ensino, buscando assim melhorar o ambiente de aprendizado dos alunos.

O aumento exponencial dos dados no decorrer dos anos através do advento da internet e de diversos dispositivos como celulares e computadores ocasionou uma revolução no que tange a gestão da informação. Essa mudança é responsável por uma geração de dados cada vez maior e mais complexa, esses conjuntos de dados são tão volumosos que softwares tradicionais de processamento de dados simplesmente não conseguem gerenciá-los. Esse grande volume de dados é o chamado *Big Data* (ORACLE, [201-?]).

O desafio para as ferramentas de *Big Data* é, entre outros, a manipulação de dados semiestruturados e não estruturados no intuito de extrair valor desses através de correlações e outros processamentos de análise e então compreendê-los para que tragam valor ao determinado meio aplicável (GALDINO, 2011).

O tratamento dos dados é realizado com o apoio de algoritmos inteligentes, que são sequências de instruções que permitem que se alcance uma conclusão e, assim, colaborar nas tomadas de decisões. Esses algoritmos podem ser considerados a “rede neural” do sistema e podem servir para fins diversos dependendo do propósito buscado por uma organização. Por exemplo, uma empresa pode compreender melhor o comportamento de um cliente, um médico pode saber se o paciente de uma clínica necessitará ser internado em determinado período ou de que maneira seria possível reduzir despesas dentro de uma empresa (GALDINO, 2011).

Os algoritmos de sistemas preditivos, que com base em dados processados “predizem” um fato com grandes probabilidades de ocorrer, são um grande desafio a ser superado nessa lacuna que existe entre aplicabilidade em tempo real, e análise de dados anteriores para se tomar decisões (GALDINO, 2011).

* 1. **Machine Learning**

*Machine Learning*, ou Aprendizado de Máquina, é um ramo da área de Inteligência Artificial que estuda sistemas capazes de aprender e criar regras a partir de dados (BERGER, 2014). Nas últimas duas décadas, o Aprendizado de Máquina se tornou um dos pilares da tecnologia da informação e, com a crescente quantidade de dados se tornando disponível, há boas razões para acreditar que a análise inteligente de dados se tornará ainda mais penetrante como fator necessário para o progresso tecnológico (SMOLA, 2010).

O principal objetivo do *Machine Learning*, é permitir que os computadores aprendam automaticamente sem intervenção humana e se ajustem de acordo com as necessidades do contexto (TURNER, 2019). O processo de aprendizado começa com dados ou observações, como instrução, experiência direta, ou exemplos de extração de padrões dos dados e uso desses padrões para fazer previsões no futuro (BOSE, 2019).

Operações de *Machine Learning* podem ser tão complexas a ponto de ser tornarem de difícil entendimento para os seres humanos. Para contornar essa dificuldade e melhorar a manutenção para seu uso em Inteligência Analítica, extrai-se regras localizadas nas complexas saídas dos algoritmos ou utiliza-se técnicas de visualização afim de compreender os dados e implementar uma solução (PARIKH, 2014). Tudo isso significa que é possível produzir de forma rápida e automática modelos que permitam analisar dados maiores e muito mais complexos, proporcionando o fornecimento de resultados mais rápidos e precisos – mesmo em uma escala muito grande (CETAX, [201-?]).

Os algoritmos de *Machine Learning* podem se enquadrar na categoria de *Supervised, Unsupervised*, ou *Reinforced Learning* (aprendizado supervisionado, não supervisionado ou por reforço, respectivamente) (TURNER, 2019).

Para o caso da aprendizagem supervisionada, espera-se que o humano forneça as entradas e as saídas desejadas e conceda o feedback com base na precisão das previsões durante o treinamento. Após a conclusão do treinamento, o mesmo algoritmo terá que aplicar que foi aplicado aos dados seguintes (TURNER, 2019). O aspecto iterativo do aprendizado de máquina é importante porque, quando os modelos são expostos a novos dados, eles são capazes de se adaptar independentemente. Eles aprendem com computações anteriores para produzir decisões e resultados confiáveis, passíveis de repetição (SAS, [201-?]).

Como nem todos os registros de dados coletados podem ser utilizados de modo a serem exemplos de aprendizado, os dados precisam ser filtrados. Cada registro de dados de treinamento é rotulado de acordo com o resultado esperado. O conjunto de dados resultante é denominado dados de treinamento. O qual é processado por uma técnica de aprendizado de máquina. Esta técnica examina a relação entre um registro de dados e a saída rotulada, e cria um modelo orientado por dados. Para qualquer novo dado, o modelo baseado em dados tenta dar o melhor resultado baseado nos dados aprendidos (JAHNKE, 2015).

O aspecto iterativo da aprendizagem supervisionada é importante pois, quando os modelos são expostos a novos dados, eles são capazes de se adaptar independentemente. Eles aprendem com computações anteriores para produzir decisões e resultados confiáveis, passíveis de repetição (SAS, [201-?]).

* 1. ***Processamento de Linguagem Natural***

O Processamento de Linguagem Natural, ou PLN, é o nome dado a técnica, que estuda como computadores interpretam a linguagem natural falada ou escrita (CHOWDHURY, 2003). Ele resulta de diversas disciplinas, incluindo ciência da computação e linguística computacional, que buscam preencher a lacuna entre a comunicação humana e o entendimento dos computadores, escalando outras tarefas relacionadas à linguagem. Por exemplo, o PLN possibilita que computadores leiam textos, ouçam e interpretem falas, identifiquem sentimentos e determinem quais trechos são importantes (SAS, [201-?]).

Existem diversos campos de uso para o PLN, como o processamento de textos, interface com usuário, reconhecimento de fala, sistemas especialistas, inteligência artificial e na interação entre *Chatbots* e seus usuários (MARTINEZ, 2010). Seus métodos têm se tornado cada vez mais sofisticados, possibilitando constantemente novas aplicações.

As máquinas de hoje podem analisar mais dados baseados em linguagem do que seres humanos, sem fadiga e de maneira consistente. Devido a massiva quantidade de dados não-estruturados que é gerada todos os dias, de registros médicos a mídias sociais, a automação será imprescindível para uma análise de texto e fala completa e eficiente (SAS, [201-?]).

* 1. ***Sistema de perguntas e respostas***

Um sistema de perguntas e respostas é caracterizado pela análise de uma pergunta previamente realizada, retornando uma ou mais respostas, baseadas na pergunta (LEE et al., 2001).

O domínio de sistema de perguntas e respostas, consiste em um domínio atrativo e um dos mais desafiadores para a área de Computação e Inteligência Artificial, pois requer a síntese de recuperação de informações, processamento de linguagem natural, representação de conhecimento, aprendizado de máquina e interfaces humano-computador (FERRUCCI et al., 2010).

O avanço nos sistemas de perguntas e respostas, pode auxiliar profissionais de diversas, seja em sistemas críticos ou sistemas tempo-sensíveis (como sistemas de saúde, comerciais, de segurança e de suporte ao consumidor) (FERRUCCI et al., 2010).

* 1. ***ChatBot***

*Chatbots* são sistemas computacionais que simulam o comportamento humano em conversas, capazes de analisar, interpretar e responder perguntas (SGANDERLA et al., 2003). São ferramentas produtivas que facilitam e agilizam as atividades entre os indivíduos, essa facilidade ocorre pela comunicação natural em que se é desenvolvida a conversa. É possível também que *chatbots* consigam aprender novos conceitos com o próprio usuário através da conversação (POLATIDIS, 2014).

Uma vez que, o *Chatbot* é um programa utilizado para conversações entre sistemas e humanos, se faz necessário uma interface gráfica apropriada para entrada e saída de dados, muitas vezes assumindo o lugar de pessoas no atendimento ao cliente. Para isso, são desenvolvidos utilizando técnicas como reconhecimento de digitação, similaridades de cadeias de caracteres ou técnicas sofisticadas de PLN (POLATIDIS, 2014).

**3 OBJETIVOS**

**3.1 Objetivos gerais**

A partir dos conceitos de *Machine Learning* e PLN, pretende-se desenvolver, para fins de aplicação do estudos propostos, um *Chatbot* que possa ser utilizado pelo IFSP Campus Votuporanga em suas redes sociais, mediando o contato com docentes, discentes ou com a comunidade externa, seja no setor administrativo, respondendo possíveis dúvidas sobre abertura de vagas, disponibilidade dos auxílios estudantis e afins, ou na biblioteca do campus, realizando consultas sobre disponibilidade de livros, horário de funcionamento, lista de serviços oferecidos e outras possíveis dúvidas.

O sistema consistirá em um *chatbot* capaz de responder as questões pré-definidas referentes ao ambiente em que está inserido. O algoritmo permitirá a implementação de novas respostas a base de informações, que podem ser consultadas pelo usuário posteriormente, o sistema ainda contará com um fluxo de simulação. O idioma utilizado na comunicação entre usuário e a plataforma será o português. Com o desenvolvimento do *chatbot* pretende-se efetivamente mostrar a relevância das tecnologias que serão estudadas por meio de sua implementação.

**3.2 Objetivos específicos**

Para a consecução deste Projeto de Pesquisa propõem-se as atividades descritas a seguir, referentes à capacitação em áreas de conhecimento relacionadas ao tema do Projeto:

* Desenvolver estudos de dados estruturados e não estruturados dentro do contexto do *Big Data*;
* Desenvolver estudos de *Machine Learning* e Inteligência Artificial;
* Estudo sobre o processamento de linguagem natural;
* Estudo sobre *chatbots*;
* Estudo da plataforma Dialogflow;
* Desenvolver um *Chatbot* que seja capaz de interagir com os usuários, realizar a análise de emoção dos mesmos e construir respostas baseadas em suas aprendizagens passadas.

**4 MATERIAIS E MÉTODOS**

Neste capítulo, apresentam-se os materiais a serem utilizados no desenvolvimento da pesquisa. O material é listado abaixo e a metodologia é descrita nos itens seguintes.

Para a criação do *Chatbot*, serão utilizadas as seguintes ferramentas: Anaconda Navigator, IDE Spyder, as bibliotecas NumPy, Scikit-learn e TensorFlow, Python e Dialogflow.

A descrição dos materiais previamente previstos para a construção do *Chatbot* pode ser observada a seguir:

* Anaconda Navigator: O Anaconda Navigator é uma interface gráfica que permite abrir aplicações, gerir pacotes, ambientes e canais sem ter de usar o terminal/janela de comando.
* IDE Spyder: O Spyder é ambiente de desenvolvimento integrado (IDE) contando com uma interface gráfica, semelhante ao Matlab, que permite a utilização da lingugem Python em um ambiente interativo, facilitando a edição de *scripts*, teste, *debugging* e visualização gráfica.
* NumPy: O NumPy é uma biblioteca Python utilizada na realização de cálculos em *Arrays* Multidimensionais. O NumPy fornece um grande conjunto de funções e operações de biblioteca que ajudam os programadores a executar facilmente cálculos numéricos.
* Scikit-learn: A scikit-learn é uma biblioteca de aprendizado de máquina de código aberto para a linguagem de programação Python. Ela inclui vários algoritmos de classificação, regressão e agrupamento incluindo máquinas de vetores de suporte (SVM), florestas aleatórias, *gradient boosting*, *k-means* e DBSCAN. É projetada para interagir com as bibliotecas Python numéricas e científicas NumPy e SciPy.
* TensorFlow: É uma biblioteca de software de código aberto para computação numérica usando grafos computacionais. Ela reúne uma série de modelos e algoritmos de *Machine Learning* e aprendizagem profunda (também conhecido como redes neurais) e os torna úteis por meio de uma metáfora comum.
* Python: É uma linguagem de programação de alto nível, de script, imperativa, orientada a objetos, interpretada e amplamente utilizada. A sua filosofia de design realça a legibilidade do código e sua sintaxe permite que os programadores desenvolvam em um número menor de linhas de código do que seria possível em linguagens como C++ ou Java. A linguagem permite construir programas claros de pequena e grande escala.
* Dialogflow: O Dialogflow é uma plataforma para comunicação orientada por diálogo e processamento de linguagem natural. Ele oferece um conjunto de desenvolvedores completo com um editor de código, uma biblioteca e várias ferramentas que podem ser usadas para criar aplicativos para o Assistente do Google ou *Chatbots*. As soluções podem ser implantadas em diferentes dispositivos e sistemas operacionais.

**4.1 Desenvolvimento de um repositório de dados**

Para a realização do treinamento supervisionado do algoritmo utilizando a IDE Spyder e as bibliotecas NumPy, Scikit-learn e TensorFlow, será desenvolvido um repositório com perguntas e respostas relacionadas ao tema de aplicação. Desta forma, o algoritmo será capaz de treinar a rede neural, a alimentando com experiências passadas. Ou seja, o *Chatbot* irá elaborar suas respostas baseadas em seu treinamento.

**4.2 Utilização do Dialogflow**

Uma outra forma de desenvolver o *Chatbot*, é mediante a utilização da plataforma Dialogflow, que de forma semelhante permite a criação de estâncias e rótulos para treinamento da rede neural. Sua utilização ocorre de forma mais orgânica e seus resultados podem ser gerados de forma mais eficiente.

**4.3 Avaliação dos resultados e aplicação nas redes sociais**

Após a criação do *Chatbot*, mediante aos dois métodos apresentados acima, serão realizados testes com usuários. Os resultados serão comparados, e o algoritmo com melhor desempenho será aprimorado e disponibilizado para uma possível utilização nas redes sociais do IFSP Campus Votuporanga.

**5. PLANO DE TRABALHO**

Para a consecução deste Projeto de Pesquisa, propõe-se o seguinte cronograma, baseado nas atividades apresentadas e propostas:

Tabela 5.1 Metas estabelecidas para a pesquisa.

|  |  |
| --- | --- |
| **METAS** | **DESCRIÇÃO** |
| **1** | Estudo dos sistemas de perguntas e respostas |
| **2** | Estudo do processamento de linguagem natural |
| **3** | Estudos de *Machine Learning* e Inteligência Artificial |
| **4** | Estudo das bibliotecas NumPy, Scikit-learn e TensorFlow |
| **5** | Desenvolvimento do Relatório Parcial - entrega até 03/07/20 |
| **6** | Estudo do Dialogflow |
| **7** | Construção do *Chatbot* |
| **8** | Realização dos testes de desempenho |
| **9** | Aplicação do *Chatbot* no ambiente das redes sociais |
| **10** | Desenvolvimento do Relatório Final - entrega até 30/11/2020 |

Tabela 5.2 Cronograma proposta para cumprimento das metas.

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | **MESES** | | | | | | | | |
| **METAS** | MAR | ABR | MAI | JUN | JUL | AGO | SET | OUT | NOV |
| **1** |  |  |  |  |  |  |  |  |  |
| **2** |  |  |  |  |  |  |  |  |  |
| **3** |  |  |  |  |  |  |  |  |  |
| **4** |  |  |  |  |  |  |  |  |  |
| **5** |  |  |  |  |  |  |  |  |  |
| **6** |  |  |  |  |  |  |  |  |  |
| **7** |  |  |  |  |  |  |  |  |  |
| **8** |  |  |  |  |  |  |  |  |  |
| **9** |  |  |  |  |  |  |  |  |  |
| **10** |  |  |  |  |  |  |  |  |  |

**6. VIABILIDADE DE EXECUÇÃO**

O projeto será executado dentro das dependências do Instituto Federal de São Paulo, no campus de Votuporanga. Poderá ser utilizada uma das salas e/ou laboratórios destinados ao desenvolvimento da programação do *Chatbot*. Também serão desenvolvidos os estudos das ferramentas de tecnologia que possibilitam a análise de dados por meio de *Machine Learning*. A criação do algoritmo e a realização dos testes será feita a partir de softwares e tecnologias gratuitas presentes e instalados nos computadores dos laboratórios de informática do próprio campus.

7. RESULTADOS ESPERADOS E DISSEMINAÇÃO

O projeto possui potencial de inovação para o IFSP Campus Votuporanga, uma vez que poderá oferecer o contato e resolução de dúvidas oriundas dos alunos, professores ou comunidade externa, de forma automatizada.

Procura iniciar o aluno participante nos estudos referentes a área de Inteligência, *Machine Learning* e NLP.

Pretende-se que os dados gerados pelo processo de construção do *Chatbot* e na formulação do repositório de dados para treinamento, sirvam como objeto de estudo e análise para outras pesquisas.

Espera-se que os resultados dessa pesquisa contribuam para o desenvolvimento de métodos, técnicas e/ou ferramentas que implementem a solução e eliminem os principais riscos identificados durante o refinamento do escopo do projeto, além da produção de artigos, relatando as principais dificuldades encontradas, as soluções propostas e os resultados parciais obtidos. O principal elemento desse trabalho será a elaboração do Relatório Final, descrevendo os principais métodos e técnicas investigadas e todos os resultados obtidos, e relatados nos artigos produzidos, além da implementação do sistema de *chatbot.*

**REFERÊNCIAS BIBLIOGRÁFICAS**

BERGER, M. L.; DOBAN, V. **Big data, advanced analytics and the future of comparative effectiveness research**. Journal of Comparative Effectiveness Research, Londres, v.3, n.2, p.167-176, 2014.

BOSE, R**. Advanced Analytics: opportunities and challenges. Industrial Management & Data Systems**, 2009.

CETAX. **Machine Learning: O que é, conceito e definição**, [201-?]. Disponível em: < https://www.cetax.com.br/blog/machine-learning/>. Acesso em 29 de out. 2018.

CHOWDHURY, G. G. **Natural language processing. Annual review of information science and technology**, Wiley Online Library, v. 37, n. 1, p. 51–89, 2003.

FERRUCCI, D.; BROWN, E.; CHU-CARROLL, J.; FAN, J.; GONDEK, D.; KALYANPUR, A. A.; LALLY, A.; MURDOCK, J. W.; NYBERG, E.; PRAGER, J. et al. **Building watson: An overview of the deepqa project**. AI magazine, v. 31, n. 3, p. 59–79, 2010.

GALDINO, Natanael. **Big Data: Ferramentas e Aplicabilidade**, 2011. Disponível em: <https://www.aedb.br/seget/arquivos/artigos16/472427.pdf>. Acesso em 29 de out. 2018.

JAHNKE, Patrick**. Machine Learning Approaches for Failure Type Detection and Predictive Maintenance**. Germany, GER: Technische Universität Darmstadt, 2015.

LEE, G. G.; SEO, J.; LEE, S.; JUNG, H.; CHO, B.-H.; LEE, C.; KWAK, B.-K.; CHA, J.; KIM, D.; AN, J. et al. **Siteq: Engineering high performance qa system using lexico-semantic pattern matching and shallow nlp**. In: TREC. [S.l.: s.n.], 2001.

MARTINEZ, A. R. **Natural language processing**. Wiley Interdisciplinary Reviews: Computational Statistics, Wiley-Blackwell, v. 2, n. 3, p. 352–357, mar 2010.

ORACLE. **O que é Big Data?**, [201-?]. Disponível em: < https://www.oracle.com/br/big-data/guide/what-is-big-data.html>. Acesso em 01 de out. 2019.

PARIKH, D. et al**. Improving rail network velocity: A Machine Learning approach to Predictive maintenance**. Transportation Research Part C: Emerging Technologies, 2014.

POLATIDIS, N. **Chatbot for admissions**. arXiv preprint arXiv:1408.6762, 2014.

SANTANCHÈ, A. **NoSQL e Big Data - Aula 27 - Bancos de Dados 2015.2**, Disponível em: <https://www.youtube.com/watch?v=-a2pyU0uhww>. Acesso em 29 de out. 2018.

SANTUCCI, Gérald. **The internet of things: between the revolution of the internet and the metamorphosis of objects**, [s.d.]. 2010.

SAS. **Machine Learning: O que é e qual sua importância?**, [201-?]. Disponível em: < https://www.sas.com/pt\_br/insights/analytics/machine-learning.html >. Acesso em 01 de out. 2019.

SGANDERLA, R. B.; FERRARI, D. N.; GEYER, C. F. R**. Bonobot: Um chatterbot para interação com usuários em um sistema tutor inteligente**. XIV Simpósio Brasileiro de Informática na Educação, 2003.

SMOLA, Alex; VISHWANATHAN, S.V.N. **Introduction to Machine Learning**. United Kingdom, UK: Cambridge University, 2008.

TURNER, R. **Python machine learning: the ultimate beginner's guide to learn python machine learning step by step using scikit-learn and tensorflow**. [S.l.: s.n.], 2019. E-book.