

# Integração entre Bases de Casos e Machine Learning: Uma proposta de *Help Desk* Inteligente

Evandro Mesales Vidi Lins e Carine Geltrudes Webber

## Resumo

Devido ao aumento de demandas relacionadas ao uso das tecnologias, novas formas de organizar e acessar o conhecimento são necessárias. Este artigo propõe utilizar uma base de dados não estruturados e heterogêneos para desenvolver um protótipo de um sistema baseado na representação de conhecimento, capaz de subsidiar a aplicação de técnicas de inteligência artificial. Isto é especialmente útil quando se deseja propor soluções automáticas para problemas reportados diariamente, oriundos do uso de dispositivos eletrônicos como computadores, impressoras, *smartphones* e *tablets*, entre outros. Mais precisamente, trata-se neste artigo, de uma central de *help desk* que atende um volume expressivo de chamados e que precisa atuar de forma eficiente e reduzir o prazo de atendimento. O estudo piloto desenvolvido e aplicado se baseou na representação do conhecimento sobre soluções para problemas corriqueiros, usualmente solucionados por colaboradores humanos em uma central de *help desk*. Nesta tarefa aplicou-se o método do Raciocínio Baseado em Casos, o processamento de linguagem natural e técnicas de *Machine Learning* na construção de um sistema piloto. Os resultados preliminares foram satisfatórios, demonstrados por meio de métricas de acurácia e satisfação do usuário.

## Palavras-chave

Raciocínio baseado em casos, *Machine learning*, Processamento de linguagem natural. Inteligência artificial.

# Case-Based Integration and Machine Learning: A Smart Help Desk Proposal

## Abstract

Due to the increasing demands related to the use of technologies, new ways of organizing and accessing knowledge are needed. This paper proposes to use an unstructured and heterogeneous database to develop a prototype of a knowledge representation based system, capable of subsidizing the application of artificial intelligence techniques. This is especially useful when you would like to propose automatic solutions to daily reported problems arising from the use of electronic devices such as computers, printers, smartphones, tablets, among others. More precisely, it relates to a help desk service that handles a significant volume of calls and that needs to act in a way to efficiently reduce the service time. The pilot study developed and applied was based on the representation of knowledge about solutions to common problems usually solved by humans from help desk. In this article, in addition to knowledge representation, we apply the Case-Based Reasoning method, natural language processing, and Machine Learning techniques to build a prototype system. Preliminary results were satisfactory, demonstrated through accuracy and user satisfaction metrics.

## Keywords

Case-based reasoning, Machine learning, Natural Language Processing. Artificial Intelligence.

Pós-Graduação em Engenharia 4.0 - Universidade de Caxias do Sul (UCS)

Emails: emvlins@ucs.br; cgwebber@ucs.br

Data de envio: 14/09/2019

Data de aceite: 18/11/2019

<http://dx.doi.org/10.18226/23185279.v7iss2p148>

## I. INTRODUÇÃO

O conhecimento é o elemento-chave das organizações. Inovações tecnológicas têm possibilitado a aquisição e análise automática de dados, produzindo conhecimento para aprimorar os processos de tomada de decisão organizacional. Se, por um lado, colaboradores detêm grande parte do conhecimento organizacional, por outro, cada vez mais as máquinas e sistemas operam, produzem e realizam tarefas além do controle humano. Esse marco tecnológico é denominado de quarta revolução industrial, também conhecido como Indústria 4.0, que surge trazendo novos conceitos e ferramentas [2].

Acompanhando a Indústria 4.0, se observa uma grande evolução tecnológica que depende da digitalização total, permeando todos os setores organizacionais, afetando processos, sistemas e equipamentos [3]. A capacidade das organizações de gerar, registrar e compartilhar conhecimento é a base para a inovação e a competitividade nas organizações [4]. Para suportar todo esse movimento utiliza-se a Inteligência Artificial (IA) que é responsável pela automação do comportamento inteligente, no ramo da ciência da computação [15].

Neste contexto, o registro adequado do conhecimento é fundamental para que se possa fazer frente aos inúmeros desafios organizacionais. Em particular, neste trabalho, entende-se que o conhecimento humano para a resolução de problemas e tomadas de decisão é crucial para o bom desempenho organizacional. Um problema recorrente nas organizações é o atendimento às demandas internas (chamados) via sistemas de *help desk*.

O aumento da demanda por serviços de *help desk* impacta em formar e manter equipes técnicas com capacidade e conhecimentos para atender as necessidades crescentes dos colaboradores. Torna-se igualmente necessário o uso de ferramentas para registro do conhecimento de forma organizada. Esse registro poderá ser utilizado para a otimização dos atendimentos não sendo tão necessário o conhecimento aprofundado de um técnico [11].

A partir de uma base de conhecimento bem estruturada é possível apoiar no processo de atendimento das demandas. Este apoio está baseado no uso de sistemas de suporte à decisão que utilizam Inteligência Artificial aplicando uma técnica de IA conhecida como Raciocínio Baseado em Casos (RBC) [5].

Esta ferramenta se utiliza de uma estrutura de casos como base para a aplicação de abordagem com o objetivo de resolver novos problemas, resgatando e adaptando soluções usadas anteriormente para resolver problemas semelhantes [6].

O presente artigo tem como objetivo propor um modelo para o registro do conhecimento. O modelo elaborado será utilizado como base para resolução de problemas aplicando técnicas de IA. Este estudo se limita a questões relacionadas a hardware justamente pela complexidade e ações necessárias para a resolução dos chamados. Como estudo futuro poderá ser estendido para chamados relacionados a outros temas.

A fim de apresentar os conceitos-chave do trabalho, o referido artigo está organizado em quatro seções. A primeira

seção discorre sobre *Conhecimento, Inteligência Artificial e RBC – Raciocínio Baseado em Casos e Mineração de Texto*. Na segunda seção apresenta-se o estudo de caso proposto, incluindo o atual cenário e o método utilizado. A terceira seção apresenta os resultados obtidos do modelo criado e, por fim, na quarta seção, a conclusão do artigo.

## I. REFERENCIAL TEÓRICO

### A. Representação do Conhecimento

A representação do conhecimento é a forma como o conhecimento sobre um determinado assunto é registrado. É descrita como o mapeamento entre os objetos e as relações de um domínio de problema [15]. A organização do conhecimento segue alguns princípios que se aplicam a uma série de domínios e podem ser suportados por uma linguagem representacional. Um exemplo são as hierarquias de classe que podem ser encontradas tanto em sistemas de classificação científicos, como em sistemas de senso comum [15].

A representação do conhecimento organiza a entrada da informação de baixo nível em um formulário estruturado que ajuda o agente a entender e resolver problemas em um determinado domínio [16].

O entendimento de representação de conhecimento e a forma como interage com a IA varia bastante. O enfoque principal é como transferir o conhecimento para um sistema computacional, permitindo que possa ter um raciocínio adequado para que o sistema interaja e se adapte ao ambiente [17].

Alguns questionamentos que refletem a complexidade e importância da representação do conhecimento: de que forma o conhecimento pode ser expresso? Como encontrar a linguagem adequada para a representação deste conhecimento? Como formar uma base de conhecimento suficientemente detalhada e que represente a compreensão do domínio? Como realizar inferências automáticas, dando acesso tanto ao conhecimento implícito, quanto aquele armazenado explicitamente? Como o sistema deve proceder na presença de informações incompletas, incorretas ou de senso comum? [17].

Com o registro do conhecimento é possível a aplicação da Inteligência Artificial (IA) para que a máquina possa aprender e posteriormente tomar decisões baseada em casos passados, simulando mecanismos da inteligência humana.

### B. Inteligência Artificial

O termo Inteligência Artificial (IA) apareceu pela primeira vez no ano de 1956 [18]. IA se refere ao ramo da ciência da computação dedicado ao desenvolvimento de algoritmos de computador para realizar tarefas tradicionalmente associadas à inteligência humana, como a capacidade de aprender e resolver problemas. IA é um termo amplo que designa uma variedade de campos e técnicas [19].

O estudo da Inteligência Artificial e seus resultados vem sendo aplicado em todos os setores da vida humana, tais como: transporte, agricultura, educação, sistemas de saúde, sistema judiciário, entre outros [10]. A IA é uma importante

tecnologia que suporta inúmeras atividades econômicas e contribui efetivamente para o crescimento sustentável de alguns países, como o Japão por exemplo. Nos últimos anos a IA tem despertado a atenção de países desenvolvidos como fator decisivo para o crescimento. A *Sciense Survey* apontou que houve aumento de 38% para 62% das empresas que utilizaram IA entre o ano de 2016 até 2018 [9].

Os sistemas baseados em conhecimento foram os primeiros a serem planejados e desenvolvidos, como ilustra a figura 1. Em um sistema baseado em conhecimento prioriza-se a construção de uma base de conhecimento, que é oriundo de entrevistas e coletas de dados com especialistas (à direita da figura 1). A partir do conhecimento, algoritmos como os usados em mecanismos de inferência, podem responder questões e consultas de usuários por meio de interfaces (à esquerda da figura 1).

A complexidade no desenvolvimento de tais sistemas está na construção das bases de conhecimento, que demanda a disponibilidade de especialistas humanos ou dados suficientemente completos e abrangentes sobre um domínio. Uma das formas de representar conhecimento especialista é através de um caso contendo uma descrição de um problema e sua solução.



Figura 1- Representação de um sistema baseado em conhecimento

### C. Raciocínio Baseado em Casos

Raciocínio Baseado em Casos (RBC) é considerado uma técnica eficaz de Inteligência Artificial, pois seu processo de raciocínio é muito próximo do processo de pensamento real no cérebro humano [14]. O RBC não é somente uma técnica aplicada em sistemas de Inteligência Artificial, mas sim uma técnica que é aplicada diariamente em decisões independente do uso de sistemas [15].

A ferramenta que utiliza o método RBC é capaz de tomar decisões apropriadas mesmo com um mínimo de informações e até com a falta das mesmas. Devido a essa capacidade a ferramenta RBC proporciona uma diminuição das intervenções de pessoas especializadas com conhecimento técnico aprofundado [11]. As técnicas de RBC permitem que o raciocínio especializado possa ser simulado, reutilizando experiências passadas [12].

O RBC pode ser representado por um ciclo de quatro etapas principais: Recuperar, Reutilizar, Revisar e Reter, conforme representado na figura 2.

- Recuperar – a partir de um problema proposto recuperar casos semelhantes para auxiliar na solução.
- Reutilizar – identificar a solução atribuída ao caso semelhante recuperado e adaptá-la para propor como solução a ser aplicada no problema atual.
- Revisar – Testar a solução adaptada para o problema atual.

- Reter – Registrar a nova solução, após ter sua validade confirmada, para que possa ser utilizada em novos casos [12], [13].

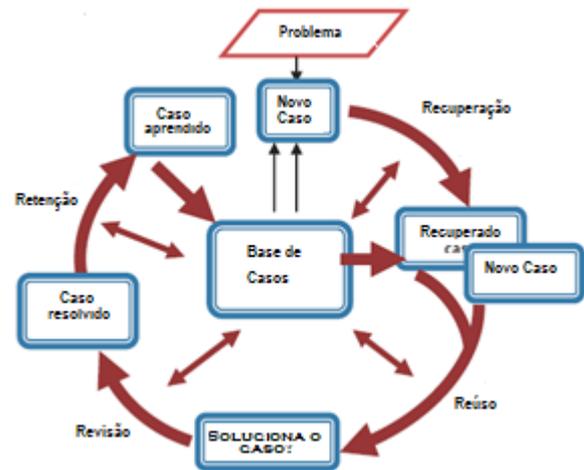


Figura 2- Representação do ciclo RBC [Adaptado de 12]

O registro do conhecimento utilizando o RBC pode contribuir significativamente para a redução nos esforços de implantação de sistemas baseados no conhecimento [12].

### D. Mineração de texto

A mineração de texto é bastante utilizada para a descoberta de conhecimento em texto e é conhecida como KDT (*Knowledge Discovered Text*). Por existir um grande número de informações registradas de forma não estruturada, entende-se que esta técnica é bem relevante e possui um alto valor comercial. O KDT utiliza abordagens de Recuperação de Informação, Processamento de Linguagem Natural e Descoberta de Conhecimento em Banco de Dados [7].

De acordo com a figura 3, as etapas para a extração de conhecimento em banco de dados são bem definidas, passando pelas etapas de seleção, processamento, transformação, mineração de dados e interpretação/avaliação [7].

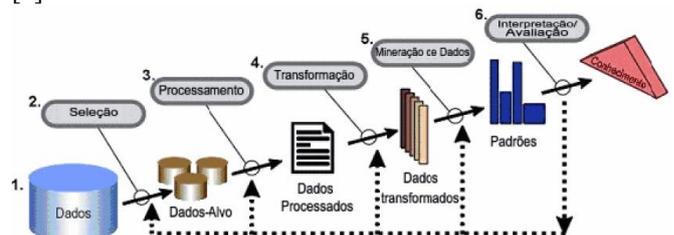


Figura 3 - Etapas KDD [7]

Já para a mineração de texto não estruturado, basicamente há duas etapas, representadas na figura 4. A primeira prevê um tratamento do texto utilizando Recuperação de Informação e Extração de Informação para torná-lo estruturado. A segunda consiste na aplicação da mineração do texto para a descoberta do conhecimento [8].

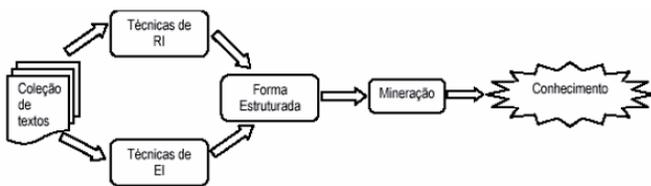


Figura 4 - Processo de mineração de textos [8]

A recuperação da informação se baseia na busca de palavras chaves ou na busca por similaridade a partir da entrada da informação pelo usuário. Posteriormente ocorre o processo de indexação atribuindo índices às palavras. O processo de indexação é composto por: Análise Léxica, Remoção de *stop-words*, *Stemming*, Seleção dos Termos Índices, Determinação de Pesos e Criação de Tesouros [7].

A Análise Léxica é responsável por separar os caracteres de entrada em palavras e separadores de palavras para posterior análise do texto. A remoção de *stop-words* realiza uma limpeza nos vocábulos removendo-os, baseada em uma lista de palavras que aparecem com muita frequência (preposições, artigos, conjunções, nomes, advérbios e adjetivos). O processo de *Stemming* é utilizado para manter somente a raiz das palavras, desprezando as variações. O conjunto de ações mencionadas neste parágrafo deve ser executado antes da indexação para reduzir o tamanho da estrutura de palavras [7].

Encontra-se na literatura metodologias de mineração de texto que propõem duas fases: a fase de refinação do texto, onde os documentos são transformados e representados em estruturas de dados; e a fase chamada destilação de conhecimento, onde padrões ou conhecimento são identificados a partir das estruturas de dados [16].

Também foram identificadas metodologias específicas com várias etapas. Na primeira se determina a finalidade da mineração do texto. Na segunda etapa as informações são coletadas, identificadas e validadas, conhecida como recuperação da informação. Nesta fase é selecionado o conteúdo relevante. Durante a terceira etapa ocorre a remoção de dados irrelevantes para o processo, através de análise léxica, tratamento e separação de palavras vazias (artigos, preposições e conjunções), tratamento flexionado de termos (os termos relacionados morfologicamente, variações de sexo, número ou tensão), tratamento de palavras compostas e normalização, obtendo-se vocábulos de raiz e rotulagem de palavras. E na quarta etapa ocorre a seleção das características desejadas, como identificar palavras-chave, identificação de entidades, indivíduos, organizações, lugares, orações, conceitos, etc. [16]

Existem diversos métodos para mineração de dados textuais. Os mais utilizados são os métodos baseados em Redes Bayesianas e Máquinas de Suporte Vetorial, por apresentarem melhores resultados em geral. Contudo, árvores de decisão, redes neurais e outras abordagens também devem ser avaliadas em cada projeto.

As linguagens de programação oferecem bibliotecas que implementam métodos variados, facilitando a realização de testes e avaliação dos desempenhos. Na linguagem Python, por exemplo, bibliotecas como *Scikit Learn*<sup>1</sup> e *TensorFlow*<sup>2</sup>

estão sendo amplamente utilizadas para análise de dados, sejam eles textuais ou estruturados. Além das bibliotecas, existem softwares, tais como: R<sup>3</sup> e Orange<sup>4</sup>, que disponibilizam algoritmos para realização de tarefas de mineração.

## II. APLICAÇÃO E MÉTODOS

O registro de problemas computacionais é uma prática seguida em várias organizações. A construção de sistemas a partir de tais registros pode ser realizada valendo-se dos sistemas de RBC. O desafio deste trabalho está em transformar um conjunto de registros textuais e heterogêneos em uma base de conhecimento esquematizada e padronizada para ser utilizada por um sistema de RBC.

A fim de desenvolver o sistema esperado entende-se que os passos a serem seguidos são os seguintes: compreender e descrever o cenário organizacional e elaborar uma proposta de desenvolvimento do software, compreendendo: criação da base de dados, definição da arquitetura de software, especificação das etapas do RBC, aplicação e avaliação dos resultados.

### A. Cenário organizacional

A empresa referência deste estudo de caso possui 20 mil funcionários distribuídos em cinco unidades fabris e quatro lojas próprias. Do volume de funcionários mencionados acima, cerca de 5 mil utilizam algum tipo de equipamento como computadores, notebooks, impressoras, leitores, *tablets* e *smartphones*, entre outros, para realização de suas tarefas diárias. Para suportar esse volume de usuários e manter a operação 24 horas em funcionamento, a empresa conta com uma central de atendimento denominada como Central de Serviços (CS). Esta equipe trabalha 24 horas, sete dias por semana e possui dois canais de atendimento, um portal onde os usuários podem abrir seu chamado e outro canal, através do telefone no ramal 600.

A empresa separa os chamados por níveis de atendimento considerando a CS como nível 0 de atendimento. Neste nível entram todos os chamados, e os que não são possíveis de solução são encaminhados para outros níveis de atendimento. Os níveis não são escalonados por complexidade e sim por especialidades. Por exemplo, o nível 1(N1) atende todos os chamados relacionados a permissões de usuários. O nível 2 (N2) é responsável por todos os chamados relacionados a hardware. Já o nível 3 se subdivide em equipe de sistemas, que atendem chamados específicos dos sistemas de ERP, RH, PLM e corporativos e, equipe de infraestrutura que responde pelos itens de comunicação, servidores, bancos de dados e segurança.

O objeto de estudo deste trabalho se restringe aos chamados que são direcionados ao N2, por ser um nicho ainda não tratado e que tem uma representatividade de 13% sobre o volume total de chamados.

1 <https://scikit-learn.org/stable/>

2 <https://www.tensorflow.org/>

3 <https://www.rstudio.com/>

4 <https://orange.biolab.si/>

**B. Criação da base de dados**

A criação da base de dados foi constituída a partir de uma primeira amostra de 257 casos, que não foi suficiente para o treinamento do classificador. Posteriormente aumentado para 433 casos, extraídos sobre um total de 10,5 mil chamados no período entre agosto de 2018 a junho de 2019. Para os 433 casos relacionados foram elencados alguns tipos de soluções. O objetivo das soluções é orientar o usuário a resolver o seu próprio problema. E, para outros casos, onde o usuário não consegue realizar o seu próprio atendimento, foram definidas soluções que abrem o chamado e o direcionam para a equipe que irá realizar o atendimento, conforme tabela com as soluções propostas.

Soluções propostas	Qtd Ocorrencias
Configurar teclado	20
Contatar TI - Enviar tecnico para alteração de layout	61
Contatar TI - Necessário reconfigurar equipamento	33
Contatar TI - Necessário Substituição.	41
Contate o setor de manutenção	24
Realizar mapeamento	10
Reinicie o equipamento	97
Retirar papel atolado	63
Revisar cabos de energia e rede	51
Seu chamado será aberto e um tecnico irá até o local	33
<b>Total Geral</b>	<b>433</b>

Tabela 1 - Soluções propostas

**C. Arquitetura de software**

Na figura 5 é apresentada a representação gráfica da arquitetura proposta para atender a necessidade do projeto em estudo. Basicamente é a forma com que o usuário entrará em contato com um atendente virtual por voz ou escrita e relatará seu problema/dificuldade/dúvida, para posterior solução.

A representação da figura 5 está proposta de uma forma macro, porém, no decorrer deste artigo será detalhada em todos os passos de acordo com o referencial teórico.



Figura 5 - Arquitetura macro

A figura 6 apresenta uma adaptação das etapas para extração de conhecimento representado na figura 3. A execução destas etapas se deu parte de forma manual e parte com a utilização da biblioteca Scykit Learn.

Neste caso, a base de dados de 433 casos é exportada do banco de dados atual da empresa em estudo para planilha

Excel. Também se realiza uma limpeza de dados, onde são removidos *stopwords* e caracteres especiais, após é aplicado o conceito de *bag of words* que consiste em realizar uma contagem de palavras com a maior frequência nos textos das descrições.

Na sequência do processo é aplicada a redução de dimensionamento fazendo a seleção e extração de características para a efetivação do treinamento e teste do algoritmo.

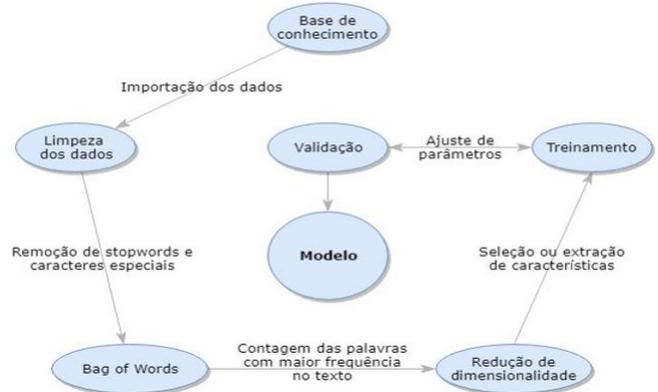


Figura 6 - Arquitetura do classificador

No processo de atendimento realizado pela empresa são geradas muitas informações que, de certa forma, acabam poluindo o texto. O atendente se utiliza de templates para coletar o maior número de informações possíveis enquanto está com o usuário ao telefone. Este procedimento baseado em perguntas é utilizado para subsidiar quem irá atender ao chamado, deixando o texto muito amplo, com perguntas e respostas.

Para a implantação de um processo utilizando técnicas de IA entende-se necessária a realização de uma limpeza no texto que será transformado em base de dados estruturado. Este passo consiste em uma das principais etapas para o sucesso do aprendizado.

Neste protótipo, o processo de limpeza dos dados foi realizado de forma manual, baseada no conhecimento de negócio. Foram retirados textos que não agregavam conhecimento para a montagem da base utilizada permitindo a sequência das próximas etapas. Na tabela 2 estão representados alguns exemplos do problema que foi descrito pelo usuário antes e depois da limpeza.

Problema antes da limpeza	Problema depois da limpeza
Usuária entrou em contato relatando que as impressões estão ficando trancadas na gaveta. Impressora XRS_DAP. Local: RH Ramal: 2270 OBS: enviado estas informações de manhã pelo Skype.	Usuária entrou em contato relatando que as impressões estão ficando trancadas na gaveta. Impressora XRS_DAP.
FABRICA/SETOR/TURNO: sala da Engenharia F4 DETALHE PROB.: Usuários da engenharia adquiriram uma balança e solicitam um técnico para realizar a instalação do software e testes com o equipamento.	Usuários da engenharia adquiriram uma balança e solicitam um técnico para realizar a instalação do software e testes com o equipamento.
Coletor não entra na janela de colocar os dados do usuario, após clicar no botão enter, o mesmo volta para a tela de inicio  Local: AlmoX Inflamaveis fab.04, proximo a TI local	Coletor não entra na janela de colocar os dados do usuario, após clicar no botão enter, o mesmo volta para a tela de inicio
FABRICA/SETOR/TURNO: Embalagem/F7 SISTEMA/PROGRAMA: - DETALHE PROB.: Computador não liga TESTES REALIZ.: - DESDE QUANDO: -	Computador não liga

Tabela 2 - Exemplos de limpeza dos textos dos problemas

Mesmo com a limpeza do texto realizada na etapa anterior, a base ainda continha pronomes, substantivos e palavras relacionadas com o próprio negócio que estão constantemente sendo mencionadas nos textos. Porém, são irrelevantes para o processo de aprendizagem.

Seguindo o fluxo da figura 6, inicia-se a remoção das *stopwords* com a criação de uma lista de palavras, contendo as mais usadas e aquelas que não contribuem para o aprendizado, conforme figura 7, que demonstra parte do código fonte do algoritmo. Nesta etapa também foram removidos os caracteres especiais. Nesta etapa do processo, e nas subsequentes foram utilizados os recursos da biblioteca Scykit.

```
stopwords = ['bom', 'dia', 'oi', 'ola', 'ola,', 'tarde', 'boa', 'noite', 'usuario', 'usuario', 'urgente',
'funcionario', 'funcionaria', 'problema', 'sala', 'equipe', 'almoxarifado',
'fabrica', 'vendas', 'reuniao', 'reunioes', 'numero', 'engenharia', 'por',
'favor', 'verificar', 'relatado', 'relatando', 'verificando', 'fab', 'portaria',
'refeitorio']

stopwords.extend(stopwords)
```

Figura 7 - Lista *Stopwords*

Nesta etapa se forma a *Bag of words*, exemplo da tabela 3, contendo as palavras relevantes ao processo de aprendizagem e a contagem das palavras com maior frequência no texto.

Destaca-se a opção em armazenar a raiz das palavras, ou seja, o elemento originário e irredutível em que se concentra a significação das palavras, para um melhor resultado do classificador.

Dado bruto	Limpeza nos dados (remoção de stopwords, acentuação e extração dos radicais)
impressora bem 40011303 não está puxando papel	impres bem 40011303 nao pux papel
O COMPUTADOR DO SIMON NAO ESTA LIGANDO	comput simon nao lig
O MICRO SEM REDE	micr red
Usuária entrou em contato relatando que as impressões estão ficando trancadas na gaveta. Impressora XRS_DAP.	entr contat relat impres esta fic tranc gavet impres xrs_dap
AO LIGAR A EMPRESSORA ESTA APARECENDO ESSE ERRO CHAMAR ASSISTENTE SC542	lig empres aparec err cham assist sc542

Tabela 3 - Exemplos formação da *Bag of words*

Para a redução de dimensionalidade não é utilizado nenhum algoritmo e sim o conhecimento de negócio para seleção dos atributos relevantes para sugerir uma solução. O atributo escolhido foi a própria descrição do problema do usuário, onde de maneira manual foi modificado o problema geral para uma descrição mais sucinta, permitindo assim que o algoritmo pudesse aprender a “raiz” do problema, conforme exemplo na tabela 3. Para a extração das características foi gerado o dataset com as palavras e quantidade de ocorrências evidenciadas na tabela 4.

Remoção de stopwords, acentuação e extração dos radicais	Extração de Características																					
	impres	nao	pux	papel	comput	simon	lig	micr	red	entr	contat	esta	fic	tranc	gavet	xrs_dap	empres	aparec	err	cham	assist	sc542
impres bem 40011303 nao pux papel	1	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
comput simon nao lig	0	1	0	0	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
micr red	0	0	0	0	0	0	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0
entr contat relat impres esta fic tranc gavet impres xrs_dap	2	0	0	0	0	0	0	0	0	1	1	1	1	1	1	1	0	0	0	0	0	0
lig empres aparec err cham assist sc542	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	1	1	1	1	1	1

Tabela 4 - Exemplos extração das características

A conclusão de um processo de *Machine Learning* passa pela etapa de submeter uma massa de dados para treinamento e teste utilizando diversos classificadores, com parâmetros exigidos de acordo com cada um dos classificadores, para se obter a melhor forma de aprendizado. Diante disso, após a preparação dos dados e a execução do treinamento e teste com alguns classificadores, os três com melhor desempenho foram o *Linear SVC*, *Random Forest* e o *Passive Agressive* mostrando uma acurácia acima de 85%. Apesar de apresentar um resultado muito próximo a outros classificadores, o *Linear SVC* entrega um tempo de treinamento e teste menor que os demais, conforme demonstrado na figura 8.

Para treinamento do algoritmo final de classificação de chamados foi utilizado o *Linear SVC* com os hiperparâmetros setados com penalidade = L2, tol = 1e-1, max\_iter=500 e dual = False. Para o restante dos hiperparâmetros foi utilizado o default do modelo que pode ser encontrado na documentação do site da *Scikit*<sup>5</sup>.

<sup>5</sup><https://scikit-learn.org/stable/modules/generated/sklearn.svm.LinearSVC.html>

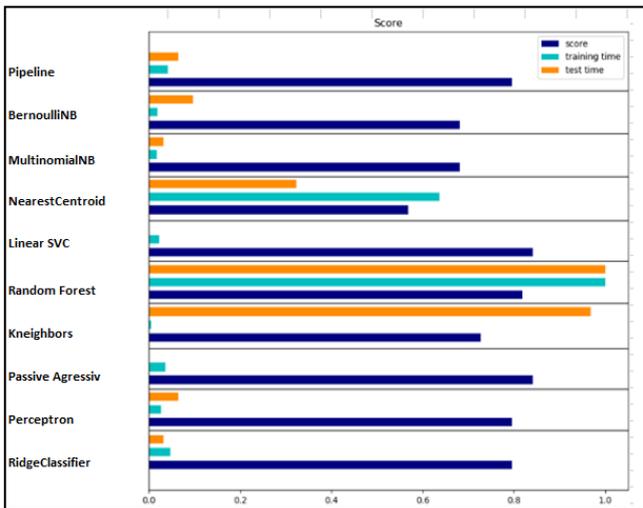


Figura 8 - Resultado dos classificadores

Na figura 9 o resultado apresentado pela matriz de confusão demonstra que o processo de treinamento e teste foi bem-sucedido, apresentando itens da classe onde o classificador conseguiu aprender e identificar as respostas corretas.

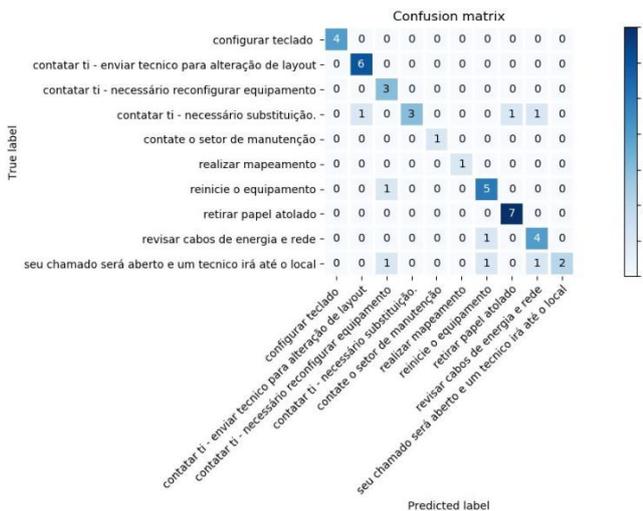


Figura 9 - Matriz de confusão

### III. RESULTADOS

O presente artigo teve como objetivo propor um modelo para o registro do conhecimento utilizando o modelo de RBC. O modelo elaborado foi desenvolvido em linguagem Python, utilizando a biblioteca ScyKit Learn. O estudo desenvolvido se limitou a questões relacionadas a problemas de hardware, totalizando 433 chamados que agrupados constituíram 10 classes.

Na análise realizada, conforme a figura 8, fica evidenciado que o classificador Linear SVC apresentou uma boa performance chegando a casa dos 85% de acurácia semelhante a outros classificadores, porém, com um tempo de treinamento e teste inferior, obtendo vantagem sobre os demais.

Com a realização do treinamento e teste do algoritmo foi desenvolvido uma interface para simular a resposta do sistema.

A figura 10 ilustra um cenário de teste do sistema

desenvolvido que basicamente extrai do texto digitado pelo usuário as características e submete ao algoritmo que atribui uma classe de solução.



Figura 10 - Interface de testes e avaliação

O sistema foi testado de forma preliminar por técnicos e programadores do grupo de desenvolvimento. Os testes foram feitos através de digitação de textos semelhantes aos utilizados pelos usuários.

Com base no material apresentado nesta seção de resultados, bem como na seção III – Aplicação e Método, pode-se afirmar que os resultados estão de acordo com a natureza do negócio avaliado e com o volume de dados analisados.

### IV. CONCLUSÃO

Ao longo das pesquisas realizadas para a construção deste artigo, fica evidente o avanço da autonomia que está sendo atribuído para máquinas e sistemas realizarem tarefas de forma inteligente. Esse avanço se dá independente de domínio e na sua maioria baseado em representações do conhecimento feito por humanos. O volume de dados e a quantidade de variáveis que influenciam os processos nas empresas só vêm aumentando. Para fazer bom uso deste conhecimento, somente com o apoio da tecnologia para conseguir transformar os dados em informações e padrões para apoiar nas decisões.

Inserido no contexto acima este artigo também busca uma forma de transformar uma base com registros em texto, não estruturados, em uma representação de conhecimento sobre a qual possam ser aplicadas técnicas de IA para auxiliar os usuários a resolver problemas do dia a dia. Desta forma, consegue-se automatizar parte do atendimento aos usuários e conseqüentemente reduzir os chamados que chegam para a Central de Serviços atender.

Ao longo do projeto realizado foi possível observar na execução das etapas para a preparação dos dados, e posterior aplicação de algoritmos de classificação, a necessidade de

uma vasta amostra de situações envolvendo problemas e soluções. Todas as etapas foram executadas com sucesso, porém, algumas de forma manual, o que pode ser um ponto de atenção para aplicação em escala.

Com relação a base de dados, pode-se concluir que os campos de título do chamado, descrição do chamado e a própria solução já atribuída pelo atendimento tradicional, são ricos em informações que podem ser utilizadas no aprendizado de máquina.

Analisando a base de dados utilizada, observou-se também que alguns problemas requerem mais de uma ação para serem totalmente solucionados. Nestes casos específicos, o chamado deve ser direcionado para técnicos específicos diminuindo assim a necessidade do envolvimento de uma pessoa da CS no atendimento da ligação. O direcionamento do chamado beneficia a CS, pois libera um atendente para auxiliar na redução do percentual de ligações perdidas que gira na casa dos 20% e gera insatisfação dos clientes. Se o total dos chamados seguir o mesmo padrão da amostra utilizada neste projeto, pode-se estimar que 40% dos chamados seriam direcionados sem passar pela CS. Estima-se uma redução de 4 mil ligações por ano ou 330 por mês.

Com o resultado do estudo piloto desenvolvido, testado e aplicado, conclui-se que é possível utilizar técnicas de Inteligência Artificial para auxiliar o atendimento dos chamados, não só os referentes a problemas de hardware como também para chamados de sistemas como o ERP, entre outros. A partir deste estudo, identifica-se que é possível estender os resultados parciais para outros setores de atendimento da informática.

Para complementar o realizado, indica-se o desenvolvimento de um assistente virtual, incorporando aspectos de interação humano-computador a fim de alcançar plenamente os objetivos esperados pela organização. Também indica-se estender o modelo aplicado aos chamados de problemas nos sistemas de ERP, PLM, RH, entre outros.

## V. BIBLIOGRAFIA

- [1] Eva Ivanová, Martin Cepel, "The Impact of Innovation Performance on the Competitiveness of the Visegrad 4 Countries," *In: Journal of Competitiveness* Vol. 10, Issue 1, pp. 54 - 72, March 2018.
- [2] Selim Erol et al., "Tangible Industry 4.0: A Scenario-Based Approach to Learning for the Future of Production," *Procedia CIRP*, vol 54, pp. 13-18, 2016.
- [3] Guilherme Salvador Corso, Ivandro Ceconello, "Uso de Realidade Aumentada como Meio de Visualização de Instruções de Trabalho," *Scientia Cum Industria*, V. 7, N. 2, PP. 94 — 101, 2019.
- [4] SANTOS DE CISNE, Caroline; HIDEEMI KANEOYA, Paula; MOURA DOS SANTOS, Luana Carla. Compartilhamento e registro de conhecimento: um estudo de caso na empresa Knowtec Sharing and recording knowledge: a case study in Knowtec (Brazil). *Revista ACB*, [S.l.], v. 20, n. 1, p. 98-111, abr. 2015. ISSN 1414-0594. Disponível em: <<https://revista.acb.org.br/racb/article/view/991>>. Acesso em: 23 jun. 2019
- [5] Diego Joel Pilger, Marcel Hugo, "Aplicação de Raciocínio Baseado em casos no suporte a decisão de um Sistema web de help desk," *Reavi – Revista Eletrônica do Alto Vale do Itajaí* n°2, 2012.
- [6] Mustafa F. Abdelwahed, Mohamed Saleh, Amr E. Mohamed, "Speeding up single-query sampling-based algorithms using case-based reasoning," *Department of Electronics, Communications and Computers*, 2018.
- [7] Eliana Cristina Nogueira Barion, Decio Lago, "Mineração de textos" *Revista de Ciências Exatas e Tecnologia* Vol.III, Nº 3, 2008.
- [8] CORREA, Adriana Cristina Giusti. "Recuperação de documentos baseada em informação semântica no ambiente AMMO". UFSCAR 2003.
- [9] Lu, H., Li, Y., Chen, M., Kim, H., & Serikawa, S. "Brain Intelligence: Go beyond Artificial Intelligence". *Mobile Networks and Applications*, 23(2), 368–375, 2017.
- [10] Hamdan O. Alanazi, Abdul Hanan Abdullah, Kashif Naseer Qureshi, "A Critical Review for Developing Accurate and Dynamic Predictive Models Using Machine Learning Methods in Medicine and Health Care," *Journal of Medical Systems*, 41:69, 2017.
- [11] Soumava Boral, Sanjay Kumar Chaturvedi, V.N.A. Naikan, "A case-based reasoning system for fault detection and isolation: a case study on complex gearboxes", *Journal of Quality in Maintenance Engineering*, Vol. 25 Issue: 2, pp.213-235, 2019.
- [12] Nadjette Dendani-Hadiby and M. Tarek Khadir, "A fault diagnosis application based on a combination case-based reasoning and ontology approach", *International Journal of Knowledge-based and Intelligent Engineering Systems*, 17, pp.305-317, 2013.
- [13] R. Schmidt and L. Gierl, Case-based reasoning for antibiotics therapy advice: an investigation of retrieval algorithms and prototypes, Elsevier *Artificial Intelligence in Medicine*, 23 pp171-186, 2001
- [14] Dongxiao Gu, Changyong Liang and Huimin Zhao, "A case-based reasoning system based on weighted heterogeneous value distance metric for breast cancer diagnosis", *Artificial Intelligence in Medicine*, Vol. 77 pp.31-47, 2017.
- [15] Luger, George F. *Inteligência Artificial: Estruturas e estratégias para a solução de problemas complexos* / George F. Luger; trad. Paulo Engel, 4.ed.- Porto Alegre: Bookmann, 2004.
- [16] NanLi, NoboruMatsuda, William W. Cohen, Kenneth R. Koedinger, "Integrating representation learning and skill learning in a human-like intelligent agent", *Artificial Intelligence* Vol.219 pp.67–91, 2015.
- [17] Brachman, Ronald J. "The future of knowledge representation". In *AAAI*, pages 1082-1092, 1990.
- [18] McCarthy J, Minsky ML, Rochester N, Shannon CE. "A proposal for the dartmouth summer research project on artificial intelligence", August 31, 1955. *AI Magazine* 27:12, 2006
- [19] An Tang, Roger Tam, Alexandre Cadrin-Chênevert, BIng, Will Guest, Jaron Chong, Joseph Barfett, Leonid Chepelev, Robyn Cairns, J. Ross Mitchell, Mark D. Cicero, Manuel Gaudreau Poudrette, Jacob L. Jaremko, Caroline Reinhold, Benoit Gallix, Bruce Gray, Raym Geis, "Canadian Association of Radiologists White Paper on Artificial Intelligence in Radiology", *Canadian Association of Radiologists Journal* 69; 120e135, 2018.