



**UNIVERSIDADE FEDERAL DO PARÁ**  
**NÚCLEO DE DESENVOLVIMENTO DA AMAZÔNIA**  
**PROGRAMA DE PÓS-GRADUAÇÃO EM COMPUTAÇÃO APLICADA**

**IGOR BRUNO LIZ GADELHA**

**O USO DE CHATBOTS NO ATENDIMENTO DE CLIENTES DE  
REVENDA POR CATÁLOGO**

*Dissertação de Mestrado*

Tucuruí  
2019



**UNIVERSIDADE FEDERAL DO PARÁ**  
**NÚCLEO DE DESENVOLVIMENTO DA AMAZÔNIA**  
**PROGRAMA DE PÓS-GRADUAÇÃO EM COMPUTAÇÃO APLICADA**

**IGOR BRUNO LIZ GADELHA**

**O USO DE CHATBOTS NO ATENDIMENTO DE CLIENTES DE  
REVENDA POR CATÁLOGO**

Texto dissertativo apresentado ao Programa de Pós-Graduação em Computação Aplicada do Núcleo de Desenvolvimento da Amazônia em Engenharia da Universidade Federal do Pará, como requisito para a obtenção do Título de Mestre em Computação Aplicada  
Área de Concentração: Ciência da Computação  
Orientador: Prof. Dr. Otávio Noura Teixeira  
Coorientador: Prof. Dra. Viviane Almeida dos Santos

Tucuruí  
2019

**IGOR BRUNO LIZ GADELHA**

**O USO DE CHATBOTS NO ATENDIMENTO DE  
CLIENTES DE REVENDA POR CATÁLOGO**

Texto dissertativo apresentado ao Programa de Pós-Graduação em Computação Aplicada do Núcleo de Desenvolvimento da Amazônia em Engenharia da Universidade Federal do Pará, como requisito para a obtenção do Título de Mestre em Computação Aplicada

Trabalho aprovado. Tucuruí, 2019:

---

Prof. Dr. Otávio Noura Teixeira  
PPCA-NDAE-UFPA

---

Prof. Dra. Viviane Almeida dos Santos  
PPCA-NDAE-UFPA

---

Prof. Dr. Heleno Fulber  
PPCA-NDAE-UFPA

---

Prof. Dr. Edson Koiti Kudo Yasojima  
UFRA

Tucuruí  
2019

Este trabalho é dedicado à Família Gadelha e ao Paranga (*in memoriam*).

## **AGRADECIMENTOS**

Agradeço aos meus pais, aos meus familiares, à minha noiva e a todos aqueles que tornaram este trabalho possível. Aos meus orientadores Prof. Dr. Otávio Noura e Prof. Dra. Viviane Almeida, por suas orientações e paciência ao longo do tempo.

## RESUMO

A presença de tecnologias que utilizam técnicas de inteligência artificial estão cada vez mais presentes no nosso dia-a-dia, o desenvolvimento de sistemas e máquinas cada vez mais capazes de realizar tarefas que outrora foram realizadas por humanos é cada vez mais presente. No setor corporativo, o atendimento de clientes por chatbots já é uma realidade e se torna cada dia mais comum. O uso de bots no atendimento de clientes, além de permitir a redução de custos, viabiliza a padronização de processos de atendimento e automação de ocorrências repetitivas. O ambiente de estudo deste trabalho foi o setor de atendimento e marketing de uma empresa especializada em revenda de produtos por catálogos. Foram identificados problemas em relação ao tempo de resposta aos clientes, recorrência de busca por atendimento fora do horário comercial, somados à elevação de custos com atendimento decorrente do crescimento do negócio. O objetivo deste trabalho de pesquisa foi implantar um chatbot capaz de auxiliar no atendimento de clientes, reduzindo as demandas repetitivas aos atendentes humanos do setor. Para garantir o alinhamento do chatbot as demandas de atendimento dos clientes, aplicou-se o método de Pareto na identificação e classificação das ocorrências do setor, direcionando-os na construção de um corpus de linguagem e modelos de componentes de processamento de linguagem natural utilizando Rasa NLU e Rasa Core.

**Palavras-chave:** Chatbots; Natural Language Processing; Rasa NLU; Rasa Core.

## ABSTRACT

Increasingly, the presence of technologies that use artificial intelligence techniques are present in our daily lives, the development of systems and machines increasingly capable of performing tasks once performed by humans is increasingly present. In the corporate sector, customer service through chatbots is already a reality and it becomes more and more common everyday. The use of bots in the customer service, besides allowing the reduction of costs, makes possible the standardization of the processes of attendance and automation of repetitive occurrences. The study environment of this work was the service and marketing sector of a company specialized in reselling products through catalogs. Problems were identified in relation to customer response time, repetitive attendance and the occurrence of out-of-hours services, in addition to the increase in costs with attendance due to the growth of the business. The objective of this research was to implement a chatbot capable of assisting customer service, reducing the repetitive demands to the human attendants of the sector. To guarantee the alignment of chatbot customer service demands, from the Pareto method, the identification and classification of the occurrences of the sector was carried out, directing in the construction of a corpus of language and models of natural language processing components using Rasa NLU and Rasa Core.

**Keywords:** Chatbot; Natural Language Processing; Rasa NLU; Rasa Core;

## LISTA DE ILUSTRAÇÕES

Figura 1 – Metodologia. ....	18
Figura 2 – Arquitetura NLP segundo (KHURANA et al.,2017) e (LIDDY,2014) - Adaptado. ....	23
Figura 3 – Arquitetura NLG. ....	28
Figura 4 – Fluxo de conversação - Fonte: (RASA,2018b). ....	36
Figura 5 – Arquitetura - Modelo 01.....	48
Figura 6 – Arquitetura - Modelo 02.....	49
Figura 7 – Arquitetura - Modelo 03.....	50
Figura 8 – Arquitetura - Modelo 04.....	51
Figura 9 – Matriz de confusão Pipeline 01.....	54
Figura 10 – Matriz de confusão Pipeline 02.....	56
Figura 11 – Matriz de confusão Pipeline 03.....	58
Figura 12 – Matriz de confusão Pipeline 04.....	60
Figura 13 – Definição do Agente – Tracy.....	61
Figura 14 – Definição das habilidades ( <i>Skills</i> ) do <i>bot</i> – Tracy.....	61
Figura 15 – Estrutura do Chatbot Proposta.....	62
Figura 16 – Fluxo Conversacional <i>Chatbot</i> .....	67
Figura 17 – Fluxo Conversacional <i>Chatbot</i> - remodelado.....	68
Figura 18 – Fluxo conversacional para expressões fora de escopo.....	69
Figura 19 – Análise do fluxo de intenções em produção.....	74
Figura 20 – Interações entre o chatbot e o cliente - Fluxo de conversa guiada.....	75
Figura 21 – Interações entre o chatbot e o cliente - Expressão fora de escopo.....	76



## LISTA DE TABELAS

Tabela 1 – Serviços NLU .....	30
Tabela 2 – Linguagens Suportadas - Fonte: (RASA,2018b).....	36
Tabela 3 – Objetivos, variáveis e referências para o método .....	40
Tabela 4 – Categorização dos atendimentos pelo canal Facebook Messenger .....	46
Tabela 5 – Intent Evaluation Results - Modelo 01 .....	53
Tabela 6 – Intent Evaluation Results - Pipeline 02 .....	54
Tabela 7 – Intent Evaluation Results - Pipeline 03 .....	55
Tabela 8 – Intent Evaluation Results - Pipeline 04 .....	56
Tabela 9 – Análise dos resultados dos classificadores de intenções .....	57
Tabela 10 – Avaliação das Usuários sobre o atendimento realizado pelo Bot .....	65
Tabela 11 – Experiência com as respostas do Bot .....	65
Tabela 12 – Expressões X Erros .....	66
Tabela 13 – Expressões dos usuários analisadas.....	72

## LISTA DE GRÁFICOS

Gráfico 1 – Publicações dos últimos 10 anos. Fonte: dimensions.ai .....	15
Gráfico 2 – Diagrama de Pareto .....	47
Gráfico 3 – Histograma Pipeline 01 .....	53
Gráfico 4 – Histograma Pipeline 02 .....	55
Gráfico 5 – Histograma Pipeline 03 .....	57
Gráfico 6 – Histograma Pipeline 04 .....	59
Gráfico 7 – Gráfico de usuários.....	70
Gráfico 8 – Formas de Interação .....	70
Gráfico 9 – Horários de Atividade do Chatbot.....	71
Gráfico 10 – Retenção na Plataforma <i>Messenger</i> .....	71

# **LISTA DE ABREVIATURAS E SIGLAS**

AI Artificial Intelligence

BOT Chatbot

CRF Conditional Random Fields

FAQ Frequency Asked Questions

IA Inteligência Artificial

MIT Massachusetts Institute of Technology

ML Machine Learning

NER Named Entity Recognition

NLP Natural Language Processing

NLU Natural Language Understanding

PLN Processamento de Linguagem Natural

SVM Support Vector Machine

VPS Virtual Private Server

# SUMÁRIO

<b>1</b>	<b>APRESENTAÇÃO .....</b>	<b>13</b>
<b>1.1</b>	<b>Introdução .....</b>	<b>13</b>
1.1.1	Contextualização.....	14
1.1.2	Mercado de Chatbots .....	14
1.1.3	Contexto científico.....	15
1.1.4	Problematização.....	15
1.1.5	Motivação .....	16
1.1.6	Justificativa.....	16
1.1.7	Objetivos.....	17
<b>1.2</b>	<b>Metodologia .....</b>	<b>17</b>
1.2.1	Métricas .....	19
<b>1.3</b>	<b>Contribuições .....</b>	<b>20</b>
<b>1.4</b>	<b>Estrutura do Trabalho .....</b>	<b>20</b>
<b>2</b>	<b>FUNDAMENTAÇÃO TEÓRICA .....</b>	<b>21</b>
<b>2.1</b>	<b><i>Natural Language Processing</i>.....</b>	<b>21</b>
2.1.1	Níveis de NLP ( <i>Levels of NLP</i> ) .....	22
2.1.2	Abordagens em NLP <i>Approachs of NLP</i> .....	26
2.1.2.1	Abordagem Simbólica <i>Symbolic Approach</i> .....	26
2.1.2.2	Abordagem Estatística <i>Statistical Approach</i> .....	26
2.1.2.3	Abordagem Conexionista <i>Conenctionist Approach</i> .....	26
<b>2.2</b>	<b><i>Natural Language Generation - NLG</i> .....</b>	<b>27</b>
2.2.1	<i>Natural Language Understanding</i> .....	29
<b>2.3</b>	<b><i>Chatbots</i>.....</b>	<b>30</b>
2.3.1	Categorização de Chatbots .....	31
2.3.2	Desafios no Desenvolvimento de Chatbots .....	32
2.3.3	Chatbots Históricos.....	33
<b>2.4</b>	<b>Rasa Stack .....</b>	<b>35</b>
2.4.1	Componentes .....	36
2.4.1.1	<i>Featurizer</i> .....	38
2.4.1.2	Classificadores .....	38
2.4.2	Avaliação de Modelos ( <i>Model Evaluate</i> ).....	38
2.4.2.1	<i>Intent Classification</i> .....	39
2.4.2.2	Entity Scoring .....	39
<b>3</b>	<b>TRABALHOS RELACIONADOS.....</b>	<b>40</b>
<b>3.1</b>	<b>Robotização nos relacionamentos: um estudo sobre o uso de <i>chatbots</i> . .</b>	<b>40</b>

3.2	<b>IBM Watson como Ambiente para Desenvolvimento e Execução de um Chatbot – Um Estudo de Caso Aplicado ao Processo de Atendimento ao Usuário</b> .....	41
3.3	<i>Enhancing Chatbots using machine learning</i> .....	43
4	<b>PROJETO</b> .....	45
4.1	<b>Coleta de Dados</b> .....	45
4.2	<b>Classificação e Construção do Diagrama de Pareto</b> .....	45
4.3	<b>Definição dos Modelos</b> .....	47
4.3.1	Critérios de Seleção .....	51
4.3.2	Testes e Validação dos Modelos .....	52
4.3.2.1	Modelo 01 .....	52
4.3.2.2	Modelo 02.....	53
4.3.2.3	Modelo 03.....	54
4.3.2.4	Modelo 04.....	55
4.3.2.5	Resultado dos Testes de Validação .....	56
4.3.3	Definições do agente.....	58
4.3.4	Modelagem de inteções do agente .....	60
4.3.5	Estrutura do Chatbot .....	62
4.3.6	Modelagem de estórias .....	62
4.4	<b>Ambientes</b> . . . . .	63
5	<b>RESULTADOS.</b> . . . . .	64
5.1	<b>Testes com usuários</b> . . . . .	64
5.2	<b>Resultados em Produção</b> . . . . .	69
5.2.1	Interações dos clientes em produção .....	71
5.2.2	Avaliação da equipe de atendimento e vendas.....	73
6	<b>CONSIDERAÇÕES FINAIS</b> .....	77
6.1	<b>Discussão</b> .....	77
6.1.1	Métodos e Resultados .....	77
6.1.2	Tecnologia, Conhecimento e Aplicabilidade.....	77
6.2	<b>Conclusão</b> .....	78
6.3	<b>Trabalhos Futuros</b> .....	79
	<b>REFERÊNCIAS</b> .....	80

# 1 APRESENTAÇÃO

## 1.1 Introdução

Os recentes investimentos realizados por grandes multinacionais no mercado de assistentes pessoais, sinalizam que existe uma maior aceitação dos clientes para interagir com a informação de uma forma mais inovadora. Os usuários deste tipo de serviço estão cada vez mais exigentes: a rapidez no acesso à informação, assim como a relevância da mesma, são aspectos determinantes neste processo (RODRIGUES,2018). Em uma era com tendências globalizantes, a informação é primordial e excessivamente veloz para qualquer tipo de negócio ou situação. A interação homem-máquina é uma alternativa disponível para que os resultados sejam alcançados de maneira rápida e precisa (MORENO et al.,2015).

O uso de *chatbots* e assistentes virtuais, tornou-se cada vez mais uma aposta das empresas para melhorar resultados no atendimento ao consumidor (OLIVEIRA et al.,2006). O atendimento ao cliente é uma área fundamental para a sua fidelização, a necessidade para as empresas dos mais diversos portes e seguimentos, fornecerem este tipo de serviço é cada vez maior e a forma mais usual de apoio ao cliente é a utilização de *Call Centers* (Centrais de Atendimento), que possuem seus custos atrelados, como despesas operacionais, recursos humanos e limitações recorrentes da necessidade de funcionamento constante (RODRIGUES,2018).

Chatbots são programas computacionais que têm como objetivo propiciar aos usuários uma interface de conversação semelhante às salas de bate-papo textual, utilizando recursos de processamento da linguagem natural e inteligência artificial, simulando um ser humano em conversação com os seus usuários (LIMA,2014). Muito antes de os termos interface conversacional ou chatbot serem inventados, Turing (1950) os descreveu como o teste final para a inteligência artificial (BRAUN et al.,2018).

Um dos principais benefícios que os *chatbots* possuem, é a sua capacidade de fornecer respostas instantâneas às perguntas, não necessitarem de descanso, férias e possuem disponibilidade de 24 horas por dia, 7 dias por semana. Acredita-se que os *bots* podem ter um grande impacto no atendimento ao cliente e abrem portas a novos modelos de atendimento (ASPECT, 2016). Sistemas baseados em *Chatbots* têm-se tornado cada vez mais populares, quando bem projetados e dotados de uma considerável base de conhecimento, alguns podem interagir com seu usuário de forma semelhante tal como um outro humano (MARCOS; MATTOS,2018).

Tanto o mercado quanto os consumidores, estão sinalizando de forma positiva o uso destas tecnologias, os relatórios sobre o mercado dos EUA emitidos por (Grand View Research,2016), (ASPECT,2016) e (ASPECT,2018), apresentam uma aceitação dos clientes de aproximadamente 40% e com grandes projeções do mercado de *chatbots* para os anos seguintes. No mercado brasileiro a situação não está muito diferente, apesar de estarmos menos maduros em comparação ao mercado norte americano e mais centralizados na região sudeste do país, Segundo (PAIVA,

2017) e (PAIVA,2018), o sucesso no mercado de assistentes virtuais, deu-se principalmente a entrada de grandes empresas de tecnologia, como Google <sup>1</sup> e Facebook <sup>2</sup>.

### 1.1.1 Contextualização

A Lucre Muito, é uma empresa do mercado amazonense com sede em Manaus e operação nos estados do Amazonas e Ceará, presente em mais de 16 cidades e em expansão em ambos os estados. A empresa é especializada na revenda de produtos por catálogo, sendo representante de grandes marcas de catálogo do país. Os produtos vendidos são das mais diversas categorias, como utensílios para o lar, cama-mesa-e-banho, roupas e maquiagens.

A empresa em sua sede possui aproximadamente 30 funcionários, divididos em 4 setores chave: marketing, atendimento, administração e logística. Durante o desenvolvimento deste trabalho, a equipe de atendimento contava com aproximadamente 5 pessoas e a equipe de marketing com 5 à 9 pessoas, aproximadamente. O processo de captação de clientes da empresa ocorre em ações nos bairros, com a realização de eventos para promover a divulgação dos produtos, contando com a participação dos representantes locais ou redistribuidores nos municípios do interior. Normalmente a primeira prospecção é realizada nestes setores, coletando dados de contato e apresentando a empresa, o produto e os canais de atendimento: telefone, email e redes sociais.

O clientes da empresa estão centrados em basicamente dois perfis: promotores e revendedores. Os promotores atuam como parceiros, realizando o cadastro de revendedores e auxiliando no processo de logística na entrega de produtos e de catálogos ao revendedores. Os revendedores são clientes diretos da empresa ou dos promotores, eles realizam a venda dos produtos da empresa presentes a partir dos catálogos impressos ou digitais aos seus respectivos clientes, recebendo um percentual sobre a venda realizada.

Com o aumento das ações, expansão da empresa e do alcance das redes sociais, houve um aumento de demandas, sendo muitas deles repetitivas para os setores de atendimento e marketing.

### 1.1.2 Mercado de Chatbots

O mapa do panorama do ecossistema brasileiro de *chatbots*, apresenta dados sobre a distribuição das empresas e do perfil dos *chatbots* desenvolvidos no Brasil. O mapa conta com a participação de 60 empresas no ano de 2017 e com 66 empresas 2018 que desenvolvem *bots* e/ou fornecem ferramentas para a sua produção no País (PAIVA,2017;PAIVA,2018).

Dentre as funções de negócio com maior expectativa de serem beneficiadas com o uso de *chatbots*, são os setores de atendimento e vendas, 95% dos entrevistados acreditam que o setor

---

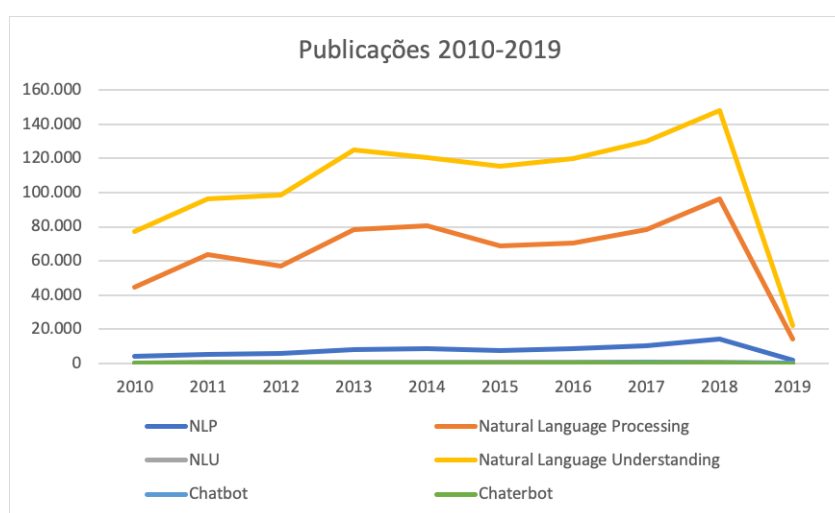
<sup>1</sup> <http://www.google.com>

<sup>2</sup> <http://www.facebook.com>

de “atendimento ao cliente” será o maior beneficiário do uso de *chatbots*, seguido do setor de vendas/marketing com aproximadamente de 53% de expectativa e 48% no processamento de pedidos.

### 1.1.3 Contexto científico

Para compreendermos o contexto científico ao qual esta pesquisa está inserido, foi realizado uma busca por palavras chaves em uma ferramenta de pesquisas de cunho acadêmico, a *dimensions.ai*<sup>3</sup>. selecionamos um espaço de 10 anos para cada palavra chave, indo de 2010 até 2019, com as seguintes palavras chave para pesquisa: “NLU”, “NLP”, “Natural Language Processing”, “Natural Language Understanding”, “Chatbot”, “Chatterbot”.



**Gráfico 1 – Publicações dos últimos 10 anos. Fonte: *dimensions.ai***

Ao analisar o gráfico1, é possível inferir que o número de citações em cada tema manteve-se crescente ao longo dos últimos 10 anos, confirmando que há alinhamento quanto ao campo de pesquisa, ao tema e ao meio científico considerando o volume de citações e publicações ao longo dos últimos dez anos das duas macro-áreas deste trabalho. Outros pontos como mercado internacional e mercado nacional de *chatbots* mostraram-se viáveis devido ao alinhamento do mercado de atuação da empresa com as expectativas de (PAIVA,2017;PAIVA,2018).

### 1.1.4 Problematização

As equipes de atendimento e marketing, perceberam que ao longo do tempo, houve uma constância de ligações e mensagens com o mesmo objetivo: orientações e dúvidas frequentes. Após o levantamento das ocorrências de atendimento, foi possível constatar o aumento por demandas repetitivas que cresciam conforme a empresa adquiria novos clientes. Os atendimentos até então não eram classificados ou categorizados, não havia um processo ou padronização ou script que guiasse no atendimento destas demandas.

<sup>3</sup> <http://dimensions.ai>



Consciente disso, buscou-se um levantamento bibliográfico para a pesquisa, com o intuito de encontrar soluções similares desenvolvidas para resolver este tipo de problema, corroborando a viabilidade do estudo e a realação com a academia e mercado.

Estudos como os de (SOUZA,2018a), (JÚNIOR; BARBOSA,2017) e (MARCOS; MATTOS,2018), utilizam chatbots, em contextos de atendimento ao cliente, buscando soluções à automação de atendimentos repetitivos, agilidade no atendimento inicial do cliente e fornecimento de informações sobre um negócio ou produto. Apesar de os trabalhos serem executados em contextos e objetos de trabalhos diferentes.

Dentro deste contexto, a partir da identificação do problema de pesquisa e identificação da similariedade de estudos, iniciou-se um estudo da efetividade do uso de tecnologias conversacionais (*Chatbots*) na resolução deste problema. O processo de pesquisa foi adequado as definições de (GIL,2002), dentre os elementos de pesquisa sugeridos pelo autor, os seguintes itens foram selecionados:

- identificação do problema;
- especificação dos objetivos;
- identificação do tipo de pesquisa;
- operacionalização das variáveis;
- elaboração dos instrumentos e determinação da estratégia de coleta de dados;
- determinação do plano de análise dos dados.

### 1.1.5 Motivação

A principal motivação do desenvolvimento é construir uma solução ao problema de atendimentos repetitivos na empresa, aplicando um método científico e tecnologia ao problema, conseguindo assim resolver total ou parcialmente o problema da empresa. Segundo (FRÖHLICH; SOARES,2018), ignorar a tendência de que a robotização de processos está virando realidade na maioria dos segmentos de atendimento a clientes pode ser fatal tanto ao empregador, quanto ao empregado que não acompanhar esta novidade tecnológica.

### 1.1.6 Justificativa

Justificamos a execução deste trabalho dado o problema real presente na empresa, a construção de um estudo e posterior solução ao problema, trazem benefícios a empresa, que pode conhecer melhor os problemas que estão atingindo ao setor a partir do método científico e com a automação, reduzir ou eliminar as demandas repetitivas que são direcionadas ao setor, podendo assim centralizar seus esforços de capital humano de atendimento ao cliente em questões de maior atenção e complexidade.

### 1.1.7 Objetivos

O objetivo geral deste trabalho é desenvolver um chatbot capaz de auxiliar no atendimento de clientes, utilizando a plataforma de mensagens *Facebook Messenger*, para atingir o objetivo de pesquisa, foram definidos como objetivos específicos:

- A classificação dos atendimentos utilizando o modelo de Pareto;
- A Definição das atividades do Bot a partir do diagrama de Pareto;
- Analisar e desenvolver modelos de conversação no atendimento de um cliente de revenda por catálogo;
- Avaliar a acurácia e satisfação obtida com o modelo proposto e posteriormente, sugerir melhorias e expansão do serviço.

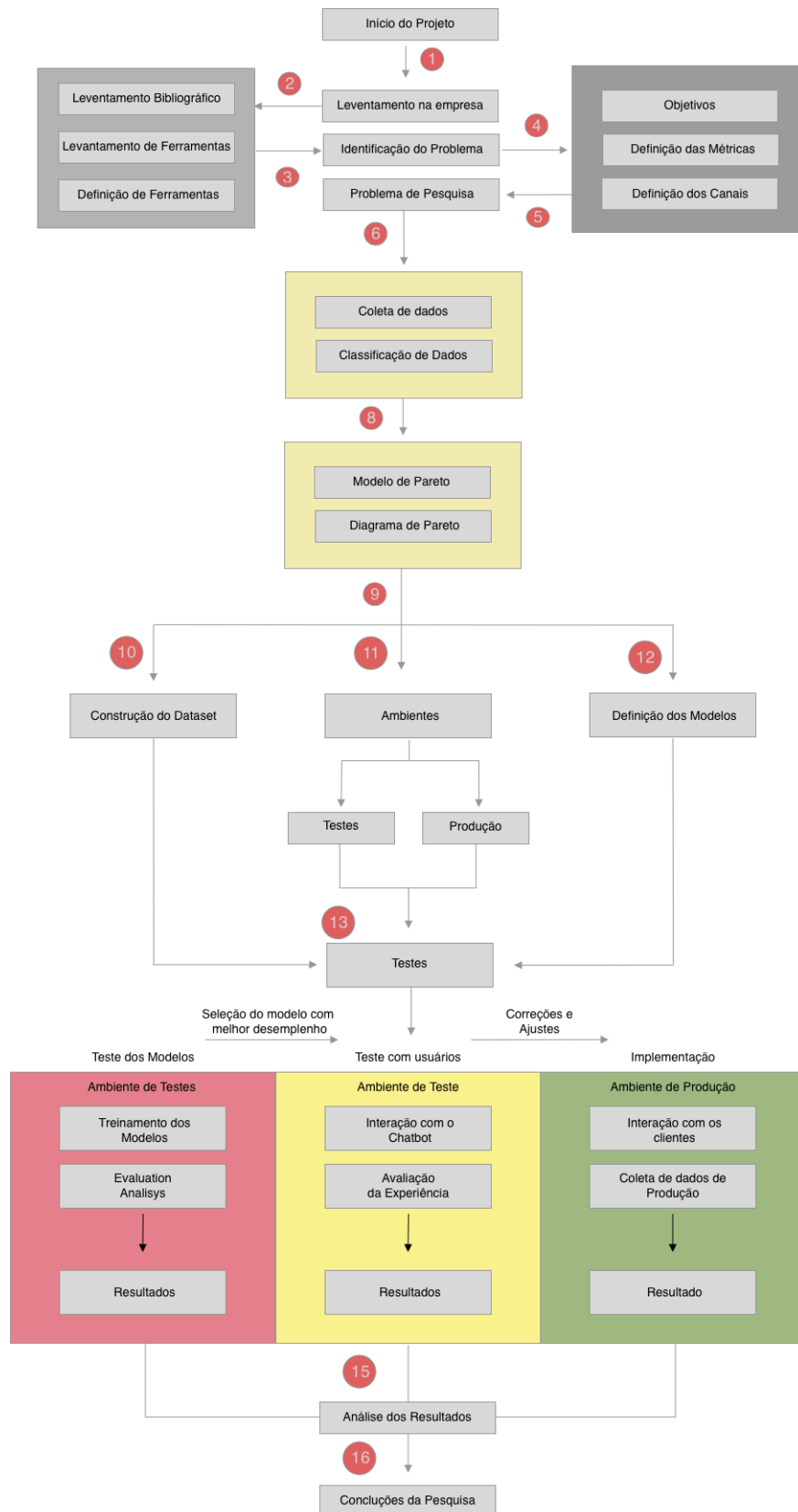
## 1.2 Metodologia

Como problema de pesquisa, o presente trabalho, buscou validar se o uso de tecnologias conversacionais, como os *chatbots*, podem auxiliar no processo de atendimento de clientes em revenda de produtos por catálogo. Para responder ao problema de pesquisa proposto, o método de pesquisa foi organizado nos seguintes momentos:

- Levantamento Teórico-técnico;
- Coleta de dados;
- Classificação e Aplicação do Modelo de Pareto;
- Definição da Ferramenta;
- Definição, testes e avaliação dos Modelos;
- Aplicação do Chatbot;
- Resultados.

A figura 1 apresenta o fluxo do processo de pesquisa do presente trabalho.

(1) Início do projeto, primeiro contato com a empresa para levantamento dos problemas de atendimento que podem ser objetos de pesquisa e o (2) levantamento bibliográfico e de ferramentas que poderiam ser utilizadas dentro daquele escopo. (3) Identificado o problema: o aumento de demandas repetitivas no setor de atendimento da empresa, (4) foram definidos os objetivos da pesquisa: Desenvolver um chatbot capaz de auxiliar no atendimento de clientes



**Figura 1 – Metodologia**

de revenda por catálogo, as métricas que seriam utilizadas para validar se os objetivos foram atingidos os canais de comunicação da empresa, formando o problema de pesquisa (5).

(6) finalizadas todas as definições, foi realizada a coleta de dados na empresa, foram coletados dados de atendimentos já realizados no setor de atendimento, após coletados, (7) os dados foram classificados de acordo com o canal de atendimento da empresa, dentre os canais disponíveis, foi selecionado o facebook messenger, por ser o canal de menor impacto direto na organização do setor, permitindo a validação e progressão posterior e por nativamente, permitir a automação por chatbots.

(8) Com os dados já classificados, foi aplicado o Princípio de Pareto, também conhecido como lei dos 80/20, diz que um pequeno número de causas (aproximadamente 20%) é responsável pela maioria dos problemas (aproximadamente 80%). Esse princípio serve de base para o Diagrama de Pareto. O diagrama de Pareto é uma forma de descrição gráfica onde busca-se identificar quais itens são responsáveis pela maior parcela dos problemas.

(9) após a classificação de dados e construção dos diagramas, o processo encaminha-se para a construção dos (10) datasets, (11) ambientes e (12) modelos. Após a etapa de definição, o método passa para a etapa de (13) testes, realizando (14) teste de acurácia dos modelos com o dataset no ambiente de testes para selecionar o modelo com melhor desempenho, teste do chatbot com usuários, utilizando o modelo com o melhor desempenho em ambiente de testes para correções e ajustes e finalmente, disponibilização em produção com os usuários finais. Os resultados finais foram analisados (15) e as (16) conclusões da pesquisa foram deliberadas ao final desta etapa.

### 1.2.1 Métricas

Como critério de seleção e validação, foram adotadas as métricas *F1-Score*<sup>4</sup>, *Precision* e *Accuracy*<sup>5</sup>, no treino e teste dos modelos, considerando os resultados de teste a de maior relevância dentre as métricas de avaliação.

Como métrica de validação, utilizamos o percentual dos atendimentos feitos pelo robô que foram transferidos aos atendimentos humanos, essa relação pode ser representado pelo total de atendimentos traferidos para um humano, que chamaremos de  $\Delta H$ , dividido pelo total de atendimentos realizados pelo *bot*, que chamaremos de  $\Delta B$ , multiplicado por 100. Esta relação pode ser expressada pela equação abaixo:

$$Atendimentos = (\Delta H / \Delta B) * 100$$

<sup>4</sup> [https://scikit-learn.org/stable/modules/generated/sklearn.metrics.f1\\_score.html](https://scikit-learn.org/stable/modules/generated/sklearn.metrics.f1_score.html)

<sup>5</sup> [https://scikit-learn.org/stable/modules/generated/sklearn.metrics.precision\\_recall\\_fscore\\_support.html](https://scikit-learn.org/stable/modules/generated/sklearn.metrics.precision_recall_fscore_support.html)

## **1.3 Contribuições**

Implementação de um método científico para classificação de problemas em uma instituição privada, o desenvolvimento de um chatbot capaz de auxiliar no atendimento de clientes, com alinhamento da construção das intenções com os problemas identificados no atendimento, somados a um corpus de linguagem construído a partir de uma amostra das interações de atendimento da empresa, sendo o produto final capaz de auxiliar para redução de demandas repetitivas empresa.

## **1.4 Estrutura do Trabalho**

O presente trabalho está organizado em 6 capítulos: capítulo de apresentação, introduzindo o ambiente de pesquisa, objetivos e metodologia; capítulo de Fundamentação Teórica, com os principais conceitos e fundamentos da pesquisa; Capítulo de trabalhos relacionados, com a apresentação de estudos similares; Capítulo de Projeto, com a discriminação das etapas e todo o processo de produção; capítulo de Resultados, apresentando aos resultados da pesquisa; e finalizando, o capítulo de Considerações Finais, apresentando as seções de Discussão, Conclusão e Trabalhos Futuros, com o comparativo dos resultados obtidos, as conclusões do autor e as sugestões de trabalhos e novos caminhos a serem tomados.

## 2 FUNDAMENTAÇÃO TEÓRICA

Estamos vivendo na era da informação e as organizações estão sendo cada vez mais transformadas, a quantidade de dados disponíveis é muito grande e em todos os segmentos organizacionais, cada vez mais é indispensável agrupar e transformar estes dados em informações relevantes para o apoio à tomada de decisões (PEREIRA; RIGO,2013).

Muito antes de os termos interfaces conversacionais ou os *chatbots* serem inventados, Turing (1950) os descreveu como o teste final para a inteligência artificial – the turing test. Apesar de sua longa história, há um recente e crescente hype sobre *chatbots* na comunidade científica e na indústria.

### 2.1 *Natural Language Processing*

Na década de 1950, a pesquisa em processamento de linguagem natural foi bastante estimulada, o campo da inteligência artificial foi profundamente revolucionado com a questão levantada por Alan Turing: "Pode uma máquina pensar?"(TURING,1950). O pensamento em si, é definido como uma habilidade dos seres humanos e de acordo com a questão levantada, Turing sugeriu o “jogo da imitação” (*The imitation Game*) como um método para evitar diretamente a questão e para especificar uma medida de realização para pesquisadores em Inteligência Artificial (SAMEERA et al.,2015). Neste período, a inteligência artificial foi descrita como a capacidade de participar de conversas escritas com uma pessoa, de modo que um revisor humano era incapaz de distinguir entre a pessoa e o computador (LACSON; KHORASANI,2011).

A *Natural Language Processing* (NLP) ou Processamento de Linguagem Natural (PNL) em português, é uma abordagem computacional para analisar textos de acordo com um conjunto de teorias, regras e tecnologias (LINDEN et al.,2003). Segundo (BRAUN et al.,2018), o NLP é um ramo da Inteligência Artificial (IA) que tem por objetivo interpretar e gerar textos em uma língua natural.

Conforme (Villmann et al.,1994), a aplicação de NLP está presente em diversas áreas do conhecimento, como computação, ciência da informação, matemática, engenharia elétrica e eletrônica, inteligência artificial, ciências cognitivas e psicologia. Já existem diversos campos de uso da NLP: processamento de textos, sumarização, interface com usuário, reconhecimento de fala, inteligência artificial, sistemas especialistas, dentre outros. Como exemplo, a NLP pode ser utilizada para realizar a interação entre os *chatbots* e seus usuários. Comprovadamente, seus métodos têm se tornado cada vez mais sofisticados e multidisciplinares (KERLY et al.,2007).

### 2.1.1 Níveis de NLP (*Levels of NLP*)

A metodologia de Níveis de Linguagem (*Levels of Language*) é utilizada para representar o processamento em um sistema NLP, composta por um modelo sincrônico de linguagem, que a partir de hipóteses nos quais os níveis de processamento da linguagem humana são suscetíveis a uma forma estritamente sequencial (LIDDY,2014). Em contrapartida, pesquisas em psicolinguísticas, apontam a um processamento mais dinâmico, nos quais estes níveis podem interagir entre si em diferentes ordens de execução.

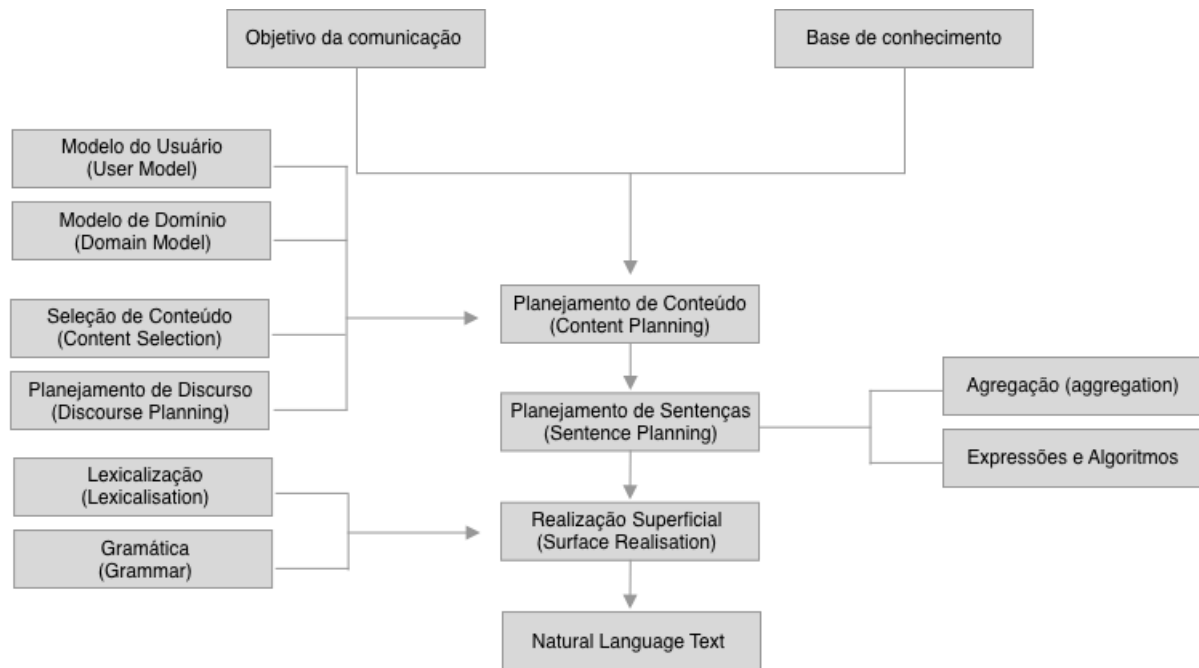
O aprimoramento da capacidade de comunicação entre humanos e robôs é uma tarefa extremamente complexa. Exige a compreensão da linguagem natural por robôs, que é uma função que deve ser realizada em todos os níveis de processamento linguístico: o nível fonético / fonológico, o morfológico, o sintático, o semântico e o pragmático (GIACHOS et al.,2017). (LACSON; KHORASANI,2011), em seu trabalho, descreve que um sistema de NLP é tipicamente composto por vários componentes, cada um ou o conjunto todos, pode ser aplicado a diferentes contextos e níveis de análise. O autor, especifica-os em:

- Conhecimento fonético e fonológico (*Phonetic and Phonologic knowledge*)
- Conhecimento morfológico (*Morphologic knowledge*)
- Conhecimento lexical (*Lexical knowledge*)
- Conhecimento sintático (*Syntatic knowledge*)
- Conhecimento semântico (*Semantic knowledge*)
- Conhecimento do discurso (*Discourse knowledge*)
- Conhecimento pragmático (*Pragmatic knowledge*)

(KHURANA et al.,2017) e (LIDDY,2014), classificam os níveis da NLP em: Fonologia (Phonology), Morfologia (Morphology), Lexical (Lexical), Sintático (Syntactic), Semântico (Semantic), Discurso (Discourse), Pragmático (Pragmatic). Apesar das nomeclaturas diferentes, os três autores abordam os mesmo níveis de estrutura. A figura2apresenta a estrutura sugerida:

#### **Conhecimentos fonético e fonológico (*Phonetic and Phonologic knowledge*)**

A fonologia é a parte da linguística que se refere ao arranjo sistemático do som. O termo fonologia vem do grego antigo e o termo fono- significa voz ou som, e o sufixo -logia refere-se a palavra ou fala (KHURANA et al.,2017). O conhecimento fonético e fonológico, refere-se à compreensão das palavras com base nos sons da fala que os produzem, como os fonemas e prosódias (LACSON; KHORASANI,2011). Este nível lida com a interpretação dos sons da fala dentro e através das palavras. Existem, na verdade, três tipos de regras usadas na análise fonológica (LIDDY,2014):



**Figura 2 – Arquitetura NLP segundo (KHURANA et al.,2017) e (LIDDY,2014) - Adaptado**

1. Regras fonéticas - para sons dentro de palavras;
2. Regras fonêmicas - para variações de pronúncia quando as palavras são faladas juntas e;
3. Regras prosódicas - para flutuação no estresse e entonação através de uma sentença

Em um sistema de NLP no qual é possível realizar entradas por meio da fala, as ondas sonoras são analisadas e codificadas em um sinal digitalizado, sendo interpretados, seguindo um conjunto de regras ou comparados com o modelo de linguagem utilizado (LIDDY,2014). Como exemplo, considere a transcrição de relatórios de radiologia, que identificam palavras de um sinal de fala. A compreensão do significado dessas palavras requer o uso de um sistema de NLP (LACSON; KHORASANI,2011).

### **Conhecimento morfológico (*Morphologic knowledge*)**

As diferentes partes da palavra representam as menores unidades de significado conhecidas como morfemas. A morfologia que compreende a natureza das palavras, é iniciada por morfemas (KHURANA et al.,2017). O Conhecimento morfológico, refere-se a forma como as palavras são construídas, nas suas unidades mais básicas (LACSON; KHORASANI,2011). Como exemplo, considere a palavra **pré-cancelamento**, a palavra possui três morfemas: um prefixo **pré**, o sufixo **mento** e raiz **cancela**. Algumas palavras não podem ser divididas, no qual o núcleo do seu significado é encontrado no seu radical, como exemplo, a palavra comer, a estas denominamos de morfema lexical (KHURANA et al.,2017).

Uma vez que o significado de cada morfema permanece o mesmo entre as palavras, um sistema NLP, a partir da sua estrutura de compreensão de linguagem, podem decifrar uma palavra



desconhecida em seus morfemas constituintes para compreender seu real significado (LIDDY, 2014).

### **Conhecimento lexical (*Lexical knowledge*)**

Na Lexical, os humanos, assim como os sistemas de NLP, interpretam o significado das palavras individuais (LIDDY,2014). Nesse nível, os seres humanos assim como os sistemas de NLP, interpretam o significado das palavras individuais (KHURANA et al.,2017). Refere-se principalmente aos significados das palavras de maneira individual, que podem ser delineadas com o sentido da palavra correspondente ou partes da fala (LACSON; KHORASANI,2011). Como exemplo, a palavra chinês pode significar uma língua (substantivo) ou um tipo de gastronomia, como em "restaurante chinês"(verbo ou adjetivo).

Dentre os diversos tipos de processamento que contribuem para o entendimento em nível de palavras, a primeira delas é a atribuição de uma única *tag* de parte da fala a cada palavra (*part-of-speech*) (LIDDY,2014). Nesse processamento, as palavras tagueadas que podem atuar como mais de uma parte da fala, são atribuídas a fala mais provável de acordo com o contexto. No nível léxico, as representações semânticas podem ser substituídas pelas palavras que têm um significado único (KHURANA et al.,2017).

### **Conhecimento sintático (*Syntactic knowledge*)**

Este nível concentra-se em analisar as palavras em uma frase de modo a descobrir a estrutura gramatical da sentença. Isso requer uma gramática e um analisador (LIDDY,2014). A saída desse nível de processamento é a representação da sentença que divulga as relações de dependência estrutural entre as palavras (KHURANA et al.,2017) e (LIDDY,2014).

Nem todas as aplicações de NLP exigem uma análise completa das sentenças, contudo, os desafios restantes na análise do escopo de definição de anexo e junção de frase preposicional não impedem mais as aplicações para as quais as dependências frasais e de cláusula são suficientes. A sintaxe em si, transmite apenas o significado na maioria das linguagens, considerando que a ordem e a dependência contribuem para a conotação (significado) (KHURANA et al.,2017) e (LIDDY,2014).

Como exemplo, considere a frase “caí, machuquei as costas, mas não quebrei nada”, o significado que pode ser extraído é que a pessoa sofreu um acidente, não houve nenhuma fratura, apesar dos machucados e possivelmente, está bem, minimizando o trauma sofrido considerando os pormenores. Diferente de “caí, não quebrei nada, mas machuquei as costas”, dando outra ênfase a oração, que apesar de não haver fratura, as costas ficaram machucadas, dando ênfase maior a lesão sofrida (LACSON; KHORASANI,2011).

### **Conhecimento semântico (*Semantic knowledge*)**

O conhecimento semântico, refere-se a como o significado de palavras individuais ao serem combinados uma sentença, adquirem um novo significado. A palavra cheiroso na frase “tá

cheiroso” pode adquirir múltiplos significados, sendo a afirmação de que a pessoa está cheirosa ou que ela está sendo inconveniente, dentro do regionalismo de linguagem paraense, denotando um valor semântico diferente (LACSON; KHORASANI,2011).

Na semântica, a maioria das pessoas pensa que o significado é determinado, no entanto, não são todos os níveis que conferem significado. O processamento semântico determina os possíveis significados de uma sentença ao articular as interações entre os significados ao nível da palavra na sentença (KHURANA et al.,2017).

Este nível de processamento pode incluir a desambiguação semântica de palavras que possuem múltiplos sentidos, de forma análoga a desambiguação sintática de palavras, contudo, podem trabalhar com múltiplas partes da fala, sendo realizada em nível sintático. A desambiguação semântica permite que um e apenas um sentido de palavras polissêmicas seja selecionado e incluído na representação semântica da sentença (LIDDY,2014). O nível semântico escrutina as palavras para a elucidação do seu dicionário e para a elucidação que deriva do meio da frase. O meio semântico que a maioria das palavras tem mais de uma elucidação, mas que podemos identificar a mais apropriada, observando o resto da frase (KHURANA et al.,2017).

### **Conhecimento do Discurso (*Discourse knowledge*)**

O conhecimento do discurso, refere-se à compreensão do texto a partir de sentenças adjacentes. Isso inclui a resolução da anáfora, ou seja, saber o que um pronome se refere a sentenças anteriores (LACSON; KHORASANI,2011). A sintaxe e a semântica funcionam com unidades de comprimento de sentença, o nível de discurso da NLP trabalha com unidades de texto maiores que uma sentença. Ou seja, não interpreta textos com múltiplas frases como apenas sentenças concatenadas, cada uma das quais pode ser interpretada individualmente. O discurso concentra-se nas propriedades do texto como um todo, que transmitem significado fazendo conexões entre sentenças componentes (LIDDY,2014).

Os dois níveis mais comuns de Resolução são: Anaphora, no qual há substituição de palavras como pronomes, que são semanticamente presos a entidades a que se referem. Reconhecimento de Discurso - O reconhecimento de estrutura de texto ou reconhecimento de estrutura do discurso, determina as funções de sentenças no texto (KHURANA et al.,2017).

### **Conhecimento Pragmático (*Pragmatic knowledge*)**

O conhecimento pragmático refere-se à compreensão de sentenças em vários contextos, muitas vezes dependendo do “conhecimento do mundo” para entender melhor o significado, além da interpretação literal (LACSON; KHORASANI,2011). O objetivo é explicar como o significado extra é lido em textos sem estar realmente codificado neles. Isso requer muito conhecimento mundial, incluindo a compreensão de intenções, planos e metas. Algumas aplicações de NLP podem utilizar bases de conhecimento e módulos de inferência (KHURANA et al.,2017).

## 2.1.2 Abordagens em NLP *Approachs of NLP*

As abordagens de processamento de linguagem natural, segundo (LIDDY,2014), estão divididas em quatro categorias: simbólica, estatística e conexionista e híbridas. (CAMBRIA; WHITE,2014) Esta seção do trabalho, apresenta os fundamentos e diferenças destas abordagens.

### 2.1.2.1 Abordagem Simbólica *Symbolic Approach*

As abordagens simbólicas, fundamentam-se na representação explícita de fatos, a partir de esquemas de representação de conhecimento e algoritmos associados, além de realizar análises profundas dos fenômenos lingüísticos presentes. Todos estes conceitos, estão fortemente atrelados aos sistemas simbólicos, regras e aos léxicos desenvolvidos pelo homem. Sistemas lógicos, baseados em regras e redes semânticas, são exemplos de abordagem simbólica (LIDDY, 2014).

Este modelo de abordagem, têm sido utilizadas há décadas, nos mais diversos campos de pesquisa e aplicações, como extração de informação, categorização de textos, resolução de ambiguidades e aquisição lexical (BASILI et al.,1996). Utilizando técnicas de aprendizado baseado em explicação, aprendizado baseado em regras, programação lógica indutiva, árvores de decisão, agrupamento conceitual e algoritmos de vizinhos mais próximos (LIDDY,2014; BASILI et al.,1996).

### 2.1.2.2 Abordagem Estatística *Statistical Approach*

Abordagens estatísticas, valêm-se do emprego de técnicas matemáticas e uso de grandes corpus de texto no desenvolvimento de modelos aproximados e generalizados de fenômenos lingüísticos. Esta abordagem é fortemente baseado em exemplos reais, fornecidos pelo corpus sem acrescentar conhecimento lingüístico ou conhecimento de mundo. Diferente das abordagens simbólicas, as abordagens estatísticas usam dados observáveis como a principal fonte de evidência (LIDDY,2014;CAMBRIA; WHITE,2014).

Esta abordagem é comumente utilizada em tarefas de reconhecimento de fala, aquisição lexical, análise, marcação de partituras, colocações, traduções automáticas estatísticas, aprendizado estatístico de gramática, dentre outras. Utiliza modelos localistas, estes modelos possuem bom desempenho em tarefas como desambiguação de sentido de palavra, geração de linguagem e inferência limitada. Modelos localistas possuem semelhanças às redes semânticas, diferindo nos links entre as unidades, que geralmente acabam não sendo rotulados (LIDDY,2014;CAMBRIA; WHITE,2014).

### 2.1.2.3 Abordagem Conexionista *Conenctionist Approach*

O conexionismo (também conhecido como “processamento paralelo distribuído”, “redes neurais” ou “neuro-computação”) tem sua origem na tentativa de projetar computadores

inspirados no cérebro (CHRISTIANSEN; CHATER,1999).

As abordagens conexionistas, utilizam modelos generalizados a partir de exemplos de fenômenos lingüísticos, de maneira similar às abordagens estatísticas. O que difere os métodos conexionistas dos métodos estatísticos, são os modelos: os modelos conexionistas combinam a aprendizagem estatística com teorias da representação. As representações conexionistas permitem a transformação, inferência e manipulação de fórmulas lógicas. Em sistemas conexionistas, os modelos lingüísticos são mais difíceis de observar devido ao fato de que as arquiteturas conexionistas são menos restritas do que as estatísticas (LIDDY,2014).

As redes conexionistas geralmente aprendem com a experiência, em vez de serem totalmente pré-especificadas por um designer. Por outro lado, modelos simbólicos de processamento de linguagem são tipicamente totalmente especificados e não aprendem (CHRISTIANSEN; CHATER,1999).

Alguns modelos conexionistas são classificados em modelos localistas e modelos distribuídos. Modelos localistas, assumem que cada unidade representa um conceito particular, modelos distribuídos são representado em função da ativação simultânea de múltiplas unidades. Esses modelos são adequados para tarefas de processamento de linguagem natural, como análise sintática, tarefas limitadas de conversão de domínio e recuperação associativa (LIDDY,2014).

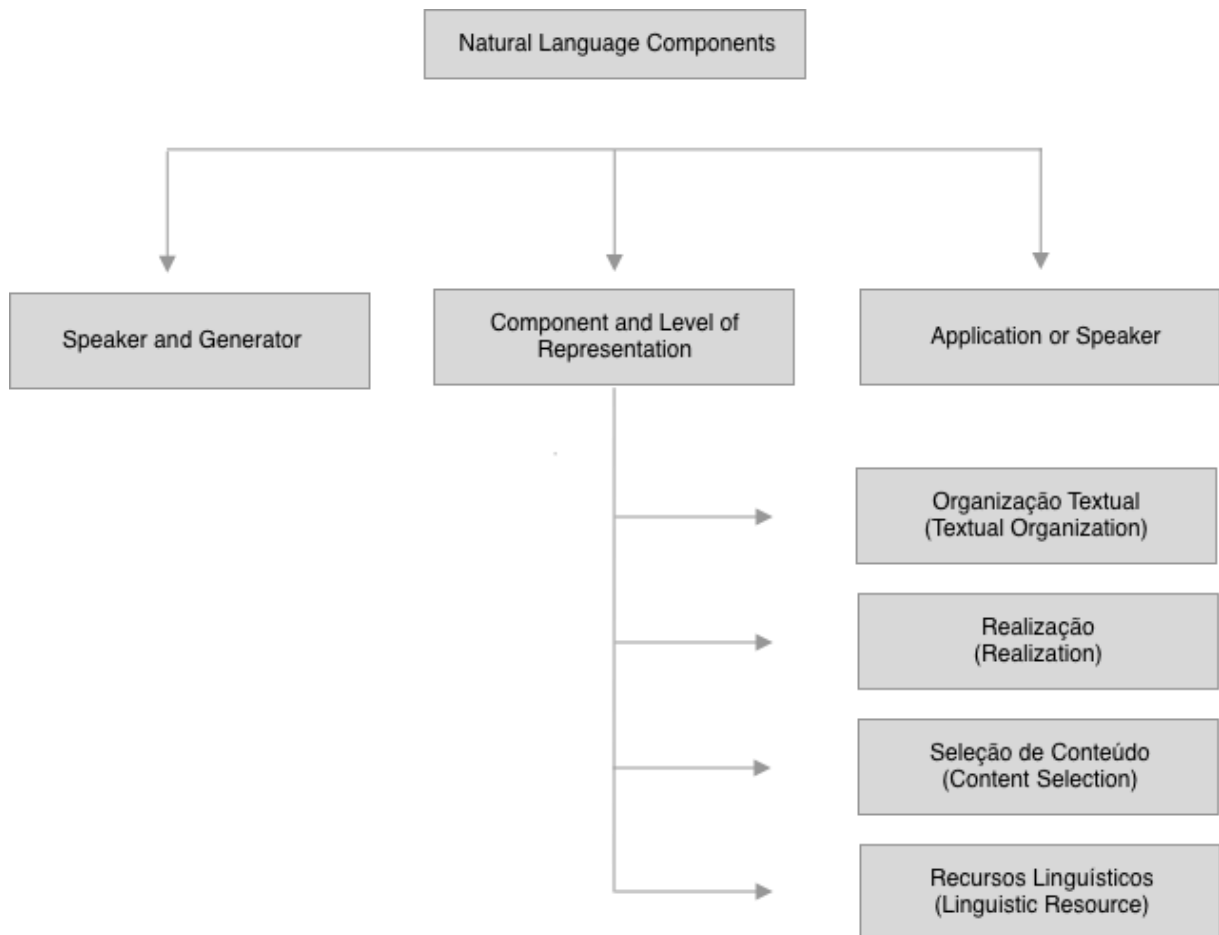
## ***2.2 Natural Language Generation - NLG***

Natural Language Generation (NLG) é uma parte da NLP e pode ser resumida como o processo de produzir frases, sentenças e parágrafos que possuem significado a parti da representação interna de um sistema (KHURANA et al.,2017). Para (KHURANA et al.,2017) e (STENT et al.,2007), o NLP acontece em três à quatro fases: identificação dos objetivos, planejamento de como os objetivos podem ser alcançados, avaliação da situação e das fontes comunicação/diálogos disponíveis e planejamento de texto.

O NLG promete portabilidade entre domínios e contextos de diálogo, a partir de conjunto de regras gerais para cada módulo gerado. Contudo, é necessário ter cautela na sua implementação, pois um NLG completo pode vir a ser demasiado lento com o uso combinado a um sistema de diálogos (STENT et al.,2007). Este estrutura modular ou componentizada, pode ser expressado de acordo com a figura3(KHURANA et al.,2017).

Os componentes de geração de linguagem natural (*Natural Language Components*), convertem uma ação abstrata de um diálogo, em expressões de linguagem natural. Um bom gerador costuma depender de vários fatores como adequação, fluência, legibilidade e variação (CHEN et al.,2017).

***Speaker and Generator*** - sua principal função é auxiliar no processa de "gerar" textos, para realizar esta ação é necessário um speaker ou uma aplicação, somados a um gerador ou um



**Figura 3 – Arquitetura NLG**

programa que transforme as intenções em frases que sejam: fluentes e relevantes para o contexto (KHURANA et al.,2017).

***Components and Levels of Representation*** - Para (KHURANA et al.,2017), O processo de geração de linguagem envolvem sub-tarefas para realização esta ação:

- Seleção de conteúdo (*Content Selection*): as informações devem ser selecionadas e incluídas no conjunto. Dependendo de como essa informação é analisada em unidades representacionais, partes das unidades podem ter que ser removidas, enquanto outras podem ser adicionadas por padrão.
- Organização textual (*Textual Organization*): A informação deve ser organizada textualmente de acordo com a gramática, sendo ordenada sequencialmente e em termos de relações linguísticas.
- Recursos linguísticos (*Linguistic Resources*): Para apoiar a realização da informação, os recursos linguísticos devem ser escolhidos. No final deste processo, esses recursos serão reduzidos a escolhas de palavras, expressões idiomáticas, construções sintáticas, etc.

- Realização (*Realization*): Os recursos selecionados e organizados devem ser apresentados como uma saída texto ou saída de voz.

*Application or Speaker* - Este componente tem função de controlar o estado atual do modelo. o Speaker dá início o processo, mas não participa da geração de idiomas. Sua responsabilidade envolve o armazenamento do histórico, estrutura o conteúdo relevante e implanta sua representação (*Realization*) de acordo com o conhecimento obtido até aquele momento (KHURANA et al.,2017).

### 2.2.1 *Natural Language Understanding*

O objetivo da NLP é realizar um processamento de linguagem natural semelhante a um ser humano. A palavra "processamento" não deve ser substituída por "compreensão", apesar de o campo da NLP tenha sido originalmente chamado de *Natural Language Understanding* (NLU) nos primórdios da IA, hoje em dia, ambas as áreas já estão distintas, a NLU é considerada uma subárea da NLP, sendo responsável pela "compreensão" de linguagem natural (LIDDY,2014).

Segundo (LIDDY,2014), para que um sistema NLU possa ser considerado completo ele deve ser capaz de:

1. Parafrapear um texto de entrada;
2. Traduzir um texto para outro idioma;
3. Responder perguntas sobre o conteúdo de um texto;
4. Inferir sobre um texto.

Para (LIDDY,2014), apesar de a NLP ter sido alvo de pesquisas ao longo dos últimos anos, principalmente com o intuito de realizar as metas relacionadas ao parafraseamento, tradução e responder perguntas sobre algum tipo de conteúdo em um texto, os sistemas de NLP não podem, por si mesmos, extrair inferências do texto, considerando assim que a NLU ainda é a meta da NLP. Existem muitas razões relacionadas ao desenvolvimento da NLU, acredita-se que três mudanças foram particularmente importantes para esse sucesso (BRAUN et al.,2018):

- Ascensão de plataformas de bate-papo universais;
- Avanços em Machine Learning (ML);
- Os avanços da Natural Language Understanding (NLU) como um serviço.

A Popularização de plataformas de conversação como o Facebook Messenger, Kik, Hangouts, WeChat, dentre outras mais, presentes no mercado, somadas aos avanços em *Machine*

Serviços	Intenções	Entidades	Importação em Lote
Luis.ai	+	+	+
Watson	+	+	+
API.ai (dialogflow)	+	+	+
wit.ai	+	+	O
Lex	+	O	-
Rasa NLU	+	+	+

**Tabela 1 – Serviços NLU**

*Learning*, nas mais diversas áreas de conhecimento, conforme (ASPECT,2018) e (Grand View Research,2016), contribuíram no processo de popularização de tecnologias conversacionais, como os chatbots (PAIVA,2018).

Os avanços e popularização de plataformas ofertando serviços em NLU, tornou mais acessível e dinâmica. A tecnologia, como as de plataformas como IBM Watson, wit.ai, api.ai (Dialogflow), dentre outras presentes no mercado, popularizaram o acesso a tecnologias NLP e NLU, oferecendo serviços de forma escalável e baseada em nuvem. Dentre as alternativas presentes no mercado e de fonte aberta, encontra-se o Rasa NLU. o Rasa NLU, que oferece a mesma funcionalidade, sem as vantagens das soluções baseadas em nuvem. Por outro lado, oferece as vantagens típicas do software de código aberto auto-hospedado (BRAUN et al.,2018; BOCKLISCH et al.,2017).

Em seu trabalho, (BRAUN et al.,2018) apresenta uma comparação das funcionalidades básicas oferecida pelos diferentes serviços presentes no mercado, durante sua pesquisa. Com exceção do Amazon Lex, todos os serviços citados compartilham os mesmos conceitos básicos: baseado em dados de exemplo, o usuário pode treinar um classificador para classificar suas intenções e entidades. A tabela é uma adaptação da tabela original publicada por (BRAUN et al.,2018), com a atualização do nome de alguns serviços.

O símbolo de + presente na tabela, indica a presença do serviço, o símbolo - a ausência do serviço e o símbolo O que o serviço está em fase de implantação ou ofertado de forma diferente. O objetivo geral dos serviços da NLU é a extração de informações semânticas estruturadas a partir da entrada da linguagem natural não estruturada. Considera-se ainda que os avanços em NLU, principalmente em serviços de extração de informação semântica estruturada a partir de entrada de linguagem natural não estruturada, são fundamentais neste processo (BRAUN et al., 2018).

## 2.3 Chatbots

Para (PEREIRA et al.,2016), um agente é um sistema com capacidade para perceber domínio e através de entradas, executar ações no ambiente ao qual está inserido (NORVIG; RUSSEL,2009). Um Chatbot, Chatterbot ou simplesmente bot, é um agente conversacional que

através do uso da linguagem natural, interage com o usuário (SHAWAR; ATWELL,2003).

O chatbot é considerado um agente conversacional que a partir de entradas em linguagem natural, torna-se capaz de realizar ações, de fornecer informações e de auxiliar usuários. Possui uma capacidade limitada de compreensão, normalmente atrelada a um domínio de conhecimento e, é capaz de interagir com um usuário através da linguagem natural, sendo capaz de decidir ou gerar uma resposta, a partir da entrada fornecida por um usuário e do contexto da conversa ao qual está inserido (RAHMAN et al.,2017).

*Chatbots* são considerados ferramentas produtivas, por facilitarem e agilizarem a realização de ações, decorrente principalmente pela comunicação natural em que é desenvolvida a interação. A indústria de *bots* ainda está em desenvolvimento, mas cresce de forma vertiginosa e em passos largos: a capacidade de compreender o que o usuário diz e ser capaz de escolher ou gerar uma resposta baseada na entrada atual e no contexto das conversas (RAHMAN et al., 2017).

Atualmente este tipo de solução vem sendo altamente explorada, permitindo que as empresas de todo mundo automatizem processos, simplificando operações com maior eficiência. Entre as principais ocorrências e, em contrapartida, as áreas que mais apresentam necessidades passíveis de atuações destes chatbot são: suporte ao atendimento ao cliente, coleta de dados, saúde, instrutor especializado, assistente pessoal e bancos, tomada de decisão em geral baseados em pergunta e resposta, por exemplo, área de atendimento e suporte, área de tutorias acadêmicas, orientações, etc (FRÖHLICH; SOARES,2018).

### 2.3.1 Categorização de Chatbots

No início da popularização dos chatbots, houveram muitas discussões sobre o que realmente seria um chatbot. Com a ascensão do processamento de linguagem natural somadas às técnicas de aprendizado de máquina, alguns dos aplicativos de conversação mais avançados e conseqüentemente, melhores recursos, para separarem-se de sua concorrência, cunharam diferentes nomenclaturas. No entanto, o mercado não se importava com o poder do robô ou com a tecnologia, desde que resolvesse os problemas. Com o tempo, muitos dos termos outrora utilizados para definir os bots, acabaram-se se tornaram sinônimos uns dos outros (MANSON, 2017). Em uma pesquisa de negócios, os chatbots podem ser agrupados nas três categorias mais comuns: *Support*, *Skills* e *Assistants*.

#### ***Support Chatbots***

Os *chatbots* de suporte (*Support Chatbots*), foram desenvolvidos para interagir em um único domínio, como o conhecimento sobre uma empresa, ou produto. Os *chatbots* de suporte precisam ter personalidade, capacidade de multi-turn e reconhecimento de contexto. Sendo capazes de orientar em qualquer processo definido no domínio e responder a uma ampla variedade de perguntas do tipo FAQ (MANSON,2017).



### ***Skills Chatbots***

Os *chatbots* com habilidades (*Skills Chatbots*), são bots do tipo single-turn que não exigem muita consciência contextual. Eles possuem comandos pré-definidos: “Ligue as luzes da minha sala de estar”, por exemplo. Eles devem ser capazes de seguir os comandos rapidamente, para que seus usuários possam executar várias tarefas ao mesmo tempo em que se envolvem com o bot (MANSON,2017).

### ***Assistant Chatbots***

Chatbots assistentes (*Assistant Chatbots*), são híbridos dos modelos citados acima. Possuem melhor funcionamento ao possuir conhecimento sobre vários tópicos. Ao construir um chatbot assistente, é importante torná-lo o mais óbvio possível sobre como o bot é treinado. A gama de perguntas que um usuário pode fazer é grande, portanto, garantir uma cobertura adequada será o fator mais difícil (MANSON,2017).

## 2.3.2 Desafios no Desenvolvimento de Chatbots

Os Chatbots possuem inúmeros desafios de desenvolvimento, para (RAHMAN et al., 2017), os principais desafios estão centrados nas limitações de construção e utilização de tecnologias e serviços, principalmente nas diferentes nuances das tecnologias que dão base ao seu desenvolvimento. Para (GURWANI,2018), existem quatro grandes desafios que as empresas irão enfrentar no desenvolvimento de chatbots em: Contextualização, atenção, testes e viabilidade de dados.

Para (RAHMAN et al.,2017), O primeiro e principal desafio do chatbot é lidar com o problemas de domínio de sintaxe em NLP. Como exemplo, se perguntarmos a um chatbot especializado em tempo e clima "qual é o tempo?", possivelmente será emitida uma resposta, entretando se expressarmos a mesma intenção da seguinte forma: "Você poderia verificar o tempo?". É possível não obtermos a resposta correta. A chave para a evolução de qualquer chatbot é sua integração com o contexto ao qual está inserido, fornecendo respostas consistentes e significativas, pois conversas sem qualquer contexto tornariam-se vagas. Torna-se um desafio para as empresas construir, desenvolver e manter a memória de bots que oferece respostas personalizadas. Partindo a um segundo conflito para IA, pois, os humanos não interagem em uma ordem definida (GURWANI,2018).

Os usuários têm um período de tempo limitado para suas consultas e esperam respostas rápidas. É um grande desafio para as empresas desenvolver chatbots, que prendem a atenção do usuário até o fim. A interface de conversação, aqui desempenha um papel importante na exibição de conversas humanas e melhores experiências do cliente. As conversas, como resultado, devem ser naturais, criativas e emocionais para que o seu chatbot seja bem sucedido. Em alguns casos, no entanto, uma máquina nem sempre geraria a mesma empatia que um ser humano poderia ter, e isso ocorre quando um substituto humano deve cuidar da solicitação do usuário (GURWANI,

2018).

O teste do Chatbot é uma questão que concentra parte da complexidade. Chatbots estão em evolução constante, principalmente em seus modelos de linguagem natural. Assim, torna-se vital testar e executar o chatbot para verificar sua precisão de compreensão, podendo ser utilizados testes automatizados e testes de lógica conversacional (manual). Para (RAHMAN et al.,2017), um chatbot em um ambiente de maior complexidade, precisa fornecer uma vasta capacidade lógica e linguística para: entradas, saídas, frases e reconhecimento de entidades. Como exemplo, considere um chatbot que realiza manipulação de consultas complexas e precisa de alta atenção no uso de formas singulares e plurais, necessita ter cuidado com sinônimos, hipónymos e realizar a análise sentimental.

O último e não menos importante ponto, (GURWANI,2018), afirma que não há necessidade de ter-se muitos dados, no preenchimento de slots inteligente ou para chatbot tecnologicamente avançado. Contudo, diferente do contexto anterior, este aspecto é vital para que o chatbot não só seja enriquecido com dados significativos, mas também esteja equipado para entregar a identidade da marca ao seu público-alvo.

### 2.3.3 Chatbots Históricos

A Artificial Linguistic Internet Computer, termo usado por (SHAWAR; ATWELL,2003) e (SHUM et al.,2018), para exemplificar um dos desafios fundamentais da inteligência artificial para conversação e compreensão em linguagem natural (NLU). Os primeiros sistemas de conversação projetados, como o (WEIZENBAUM,1966), foram projetados para imitar o comportamento humano em uma conversação baseada em texto, realizando ações específicas e delimitadas dentro de um escopo controlado. Tais programas eram frequentemente projetado para convincentemente simular como um ser humano se comportaria como um parceiro conversacional (SHUM et al.,2018). Apesar grande sucesso, a maioria destes sistemas eram principalmente baseados em regras e funcionaram bem apenas em ambientes restritos (SHAWAR; ATWELL, 2003).

#### **Eliza:**

O ELIZA é um programa que opera dentro do sistema de compartilhamento de tempo do MAC no MIT <sup>1</sup>, o que possibilita certos tipos de conversação em linguagem natural entre o homem e o computador. Quando em conversa com ELIZA, o usuário digita alguma declaração ou conjunto de instruções, usando as estruturas normais de pontuação e sentenças, exceto o ponto de interrogação, pois é interpretado como um caractere de exclusão de linha pelo sistema MAC. As sentenças de entrada são analisadas com base nas regras de decomposição, que são acionadas por palavras-chave que aparecem no texto de entrada. As respostas são geradas pelas regras de remontagem associadas às regras de decomposição selecionadas (WEIZENBAUM,1966).

---

<sup>1</sup> <http://www.mit.edu>

**Alice:**

ALICE é um robô de software ou programa ou chatbot com conhecimento do conhecimento sobre padrões de conversação em inglês. AIML, ou Artificial Intelligence Mark-up Language, é um derivado da Extensible Mark-up Language (XML). Foi desenvolvido pela comunidade de software livre Alicebic da Alic durante 1995-2000 para permitir que as pessoas introduzissem conhecimento de padrões de diálogo em chatbots baseados no software livre A.L.I.C.E (SHAWAR; ATWELL,2003).

**Elizabeth:**

Elizabeth é uma adaptação do chatbot Eliza, no qual os vários mecanismos de seleção, substituição e armazenamento de frase foram aprimorados e generalizados para aumentar a flexibilidade e a adaptabilidade (potencial). O conhecimento é armazenado como um script em um arquivo de texto, onde cada linha é iniciada com uma notação de comando de script (WEIZENBAUM,1966).

A grosso modo, o procedimento do software é bastante simples: o texto é lido e inspecionado quanto à presença de uma palavra-chave. Priscila Aliás O procedimento grosseiro do programa é bastante simples; o texto é lido e inspecionado quanto à presença de uma palavra-chave. Se tal palavra for encontrada, a sentença é transformada de acordo com uma regra associada à palavra-chave, se não uma observação livre de conteúdo ou, sob certas condições, uma transformação anterior é recuperada. O texto assim computado ou recuperado é então impresso. Se palavra for encontrada, a sentença é transformada de acordo com uma regra associada à palavra-chave, se não uma observação livre de conteúdo ou, sob certas condições, uma transformação anterior é recuperada. O texto assim computado ou recuperado é então impresso (WEIZENBAUM,1966).

**XiaoIce:**

XiaoIce, literalmente o “Little Bing”, foi projetado como uma personalidade feminina, com 19 anos de idade, fortes habilidades de linguagem, percepção visual, dentre outras 180 habilidades implementadas. Atualmente, a XiaoIce possui mais de 100 milhões de usuários únicos em todo o mundo e obteve mais de 30 bilhões de interações com humanos (SHUM et al., 2018).

O chatbot possui um Dialogue Manager (DM), encaminhando corretamente as entradas dos usuários aos respectivos módulos responsáveis. Considerando diferentes cenários, o DM coordena os módulos, a fim de gerar uma saída adequada ao contexto da conversa (SHUM et al., 2018).

Nos três anos desde que o XiaoIce foi lançado em plataformas sociais como WeChat e Weibo na China, ela se tornou uma celebridade da Internet, aparecendo como âncora meteorológica e de notícias, hospedando programas de TV e trabalhando como repórter de jornal (SHUM et al.,2018).

## 2.4 Rasa Stack

Os sistemas de conversação estão tornando-se cada vez mais difundidos como base de interação entre computadores e humanos, a medida que buscamos maneiras mais naturais de integrar a automação em nossa vida cotidiana. Exemplos bem conhecidos de conversação utilizando inteligência artificial são a Siri da Apple, Alexa da Amazon e a Cortana da Microsoft (BOCKLISCH et al.,2017).

Bibliotecas modernas de código aberto são mantidas com um alto padrão de profissionalismo, e isso se estende a implementações de algoritmos de aprendizado de máquina. Há uma grande quantidade de trabalho não relacionado à pesquisa envolvido na manutenção de um projeto amplamente utilizado, e o código produzido por grupos de pesquisa muitas vezes fica aquém das expectativas de mercado (BOCKLISCH et al.,2017).

O Rasa Stack é um *toolkit* de aprendizado de máquina de *open source* que permite que os chatbots possam expandir suas capacidades de conversação utilizando ML. O toolkit é dividido em duas ferramentas: O Rasa NLU e o Rasa Core: Rasa NLU é uma biblioteca para NLU, capaz de realizar classificação de intenções e extração de entidades, Rasa Core é um framework para gerenciamento de diálogos utilizando machine learning (RASA,2018b).

### Arquitetura:

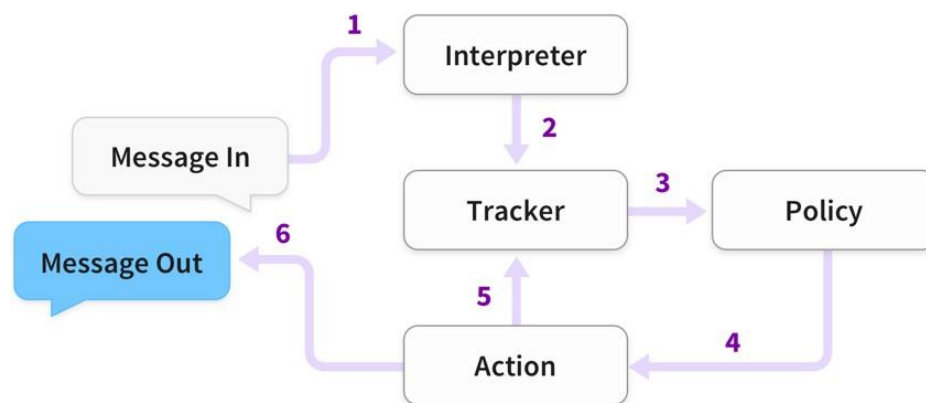
O padrão de arquitetura Rasa (Core + NLU) é fundamentalmente modular, com alto desacoplamento e de fácil integração com outros sistemas e serviços. Como exemplo, o Rasa Core pode ser usado como gerenciador de diálogos em conjunto com serviços NLU, como Dialogflow ou IMB Watson, funcionando de forma totalmente independente do Rasa NLU. O Rasa NLU, também pode ser usado de forma independente, com outros sistemas de gestão de diálogos, como o Botman <sup>2</sup> ou o Botkit <sup>3</sup> (BOCKLISCH et al.,2017).

Ambas as ferramentas possuem o código implementado em linguagem Python, além de ambas serem expostas por API's HTTP, tornando a sua integração a outros projetos, outras linguagens de programação e ferramentas, relativamente fácil (RASA,2018b). A Figura4, apresenta o fluxo de conversação dentro da arquitetura Rasa. O passo 01 é realizado pelo Rasa NLU e todos os demais passos são realizados pelo Rasa Core.

Quando uma mensagem é recebida (Message In), ela é passada para um Interpretador (*Interpreter*) (1), o interpretador irá extrair a intenção (*Intent*), as entidades (*Entities*) e qualquer outra informação estruturada que possa estar disponível. Concluindo esta etapa, o processo é repassado para o tracker (2), que mantém o estado da conversa naquele ponto, até ser notificado de que uma nova mensagem foi recebida. Finalizado este processo, O Policy recebe o estado atual do tracker (3) e decide qual melhor ação (*Action*) a ser executada (4). Após a escolha da ação, A Action escolhida é registrada pelo tracker e executada (5), o pode incluir o envio de uma

<sup>2</sup> <http://botman.io>

<sup>3</sup> <http://botkit.ai>



**Figura 4 – Fluxo de conversação - Fonte: (RASA,2018b)**

mensagem para o usuário (Message Out) (6), caso a ação executada (prevista) não for de Listen (ouvir), ele deverá retornar para a etapa 3 (BOCKLISCH et al.,2017).

O mecanismo de diálogo do Rasa Core, utiliza rastreadores (*trackers*) e estados (*states*), criando um objeto de acompanhamento por sessão de conversação para cada usuário. Os *trackers*, utilizam de um segundo recurso: os *slots* (RASA,2018b). *Slots* são como “espaços” que podem ser usados para armazenamento de informações, de forma fixa ou temporária, os tracker os utilizam para para guardarem todos os eventos (mensagens, respostas e estados) em um diálogo (conversa). A principal finalidade deste recurso, é deixar fácil o rastreamento em cada ponto de um diálogo, podendo ser reconstruído e reproduzindo todos os eventos ocorridos (BOCKLISCH et al.,2017).

O Rasa NLU, oferece suporte aos principais idiomas, estando restrito apenas ao backend escolhido, cada backend possui suas vantagens e desvantagens de uso, oferecendo um método próprio para adicionar novas linguagens, por padrão as línguas inglesa, alemã, espanhola, portuguesa, italiana, holandesa, francesa e chinesa são suportadas. A tabela2, apresenta a relação dos componentes com as linguagens suportadas.

backend	Línguas suportadas
spacy-sklearn	inglês (en), alemão (de), espanhol (es), português (pt), italiano (it), holandês (nl), francês (fr)
MITIE	inglês (en)
Jieba-MITIE	chinês (zh)
Tensorflow_embedding	Suporte a qualquer linguagem.

**Tabela 2 – Linguagens Suportadas - Fonte: (RASA,2018b)**

### 2.4.1 Componentes

#### Inicializadores:

- **NLP\_SPACY** - Inicializa estruturas SpaCy. Todos os componentes do spacy dependem disso, portanto, isso deve ser colocado no início de cada pipeline que usa qualquer compo-

nente spacy.

- **NLP\_MITIE** - Inicializa estruturas Mitie. Todos os componentes do Mitie dependem disso, portanto, isso deve ser colocado no início de cada pipeline que usa qualquer componente Mitie.

### ***Named Entity Recognition - NER***

**NER\_CRF** - O componente é aconselhável para uso em *chatbots* ou aplicativos que utilizam interação por voz ou entidades personalizadas (*custom entities*). Devido a sua habilidade de extrair entidades nomeadas dentro de cada contexto de domínio especificado, independente do idioma utilizado. Como exemplo, considere um *bot* de um restaurante que irá receber os pedidos dos clientes, neste caso específico a palavra “chinês”, pode ser identificada como culinária, mas em um aplicativo de aprendizagem de idiomas, significaria um idioma, diferindo completamente do contexto anterior.

**NER\_SPACY** - Este componente é capaz de extrair entidades já presentes nos modelos da biblioteca SpaCy, não sendo recomendado seu treinamento utilizando a biblioteca, devido a existência de vetores pré-treinados e modelos, ambos são altamente sensíveis.

**NER\_DUCKLING\_HTTP** - é um módulo da biblioteca Duckling, desenvolvida utilizando linguagem *Cloujure*<sup>4</sup>. O módulo é capaz de analisar textos e convertê-los em dados estruturados, especialmente expressões de séries temporais em inglês, espanhol, francês, italiano e chinês. O módulo é capaz de retornar de forma estruturada expressões que são repassadas em linguagem natural, dependendo apenas dos exemplos adicionados no corpus de treinamento.

**NER\_MITIE** - Utiliza o modelo de extração MITIE para localizar entidades em uma mensagem. O classificador subjacente está usando um SVM (*Support Vector Machine*) linear de várias classes com um kernel linear esparsos e recursos personalizados. O componente MITIE não fornece valores de confiança da entidade.

**NER\_CRF** - Este componente implementa **Conditional Random Fields** (CRF) para fazer o reconhecimento da entidade nomeada (**Named Entity Recognition** NER). Os CRFs podem ser considerados como uma cadeia de Markov não direcionada, onde os passos de tempo são palavras e os estados são classes de entidade. Recursos das palavras (letras maiúsculas, marcação POS, etc.) fornecem probabilidades para certas classes de entidade, assim como as transições entre tags de entidades vizinhas: o conjunto mais provável de tags é então calculado e retornado.

**NER\_SYNONYMS** - Se os dados de treinamento contiverem sinônimos definidos (usando o atributo *value* nos exemplos de entidade), esse componente garantirá que os valores da entidade detectada sejam mapeados para o mesmo valor.

### **Tokenizadores:**

---

<sup>4</sup> <https://clojure.org/>

**TOKENIZER\_WHITESPACE** - Cria um token para cada sequência de caracteres separada por espaço em branco. Pode ser usado para definir tokens para o extrator de entidade MITIE.

**TOKENIZER\_MITIE** - Cria tokens utilizando o MITIE Tokenizer. Pode ser utilizado para definir tokens para as entidades extraídas pelo Mitie.

**TOKENIZER\_SPACY** - Cria tokens utilizando o MITIE Tokenizer. Pode ser utilizado para definir tokens para as entidades extraídas pelo Mitie.

#### 2.4.1.1 Featurizer

**INTENT\_FEATURIZER\_MITIE** - Cria feature para as intenções classificadas a partir do MITIE featurizer.

**INTENT\_FEATURIZER\_SPACY** - Cria feature para as intenções classificadas a partir do MITIE featurizer.

**INTENT\_FEATURIZER\_COUNT\_VECTORS** - Cria uma representação de recursos de intenção com o uso de palavraschave usando o CountVectorizer da sklearn. Todos os tokens que consistem apenas de dígitos (por exemplo, 123 e 99) serão atribuídos à mesma característica.

#### 2.4.1.2 Classificadores

**INTENT\_CLASSIFIER\_TENSORFLOW\_EMBEDDING** - O classificador de intenções tensorflow embedding, incorpora entradas de usuário e rótulos de intenção no mesmo espaço. Utiliza aprendizado supervisionado, tendo seus modelos treinados maximizando a similaridade entre eles. O classificador precisa ser precedido por um featurizer na pipeline. Este featurizer cria os recursos utilizados. Recomenda-se usar intent\_featurizer\_count\_vectors que pode ser opcionalmente precedido por nlp\_spacy e tokenizer\_spacy.

**INTENT\_CLASSIFIER\_MITIE** - Esse classificador utiliza a biblioteca MITIE para executar a classificação de intenção. O classificador subjacente está usando um SVM linear de várias classes.

**INTENT\_CLASSIFIER\_SKLEARN** - O classificador de intenção sklearn treina um SVM linear que é otimizado usando uma grid search. Além de outros classificadores, também fornece classificações dos rótulos (tags) que não “vencem”. O classificador de intenção do spacy precisa ser precedido por um featurizer no pipeline. Esse featurizer cria os recursos usados para a classificação.

#### 2.4.2 Avaliação de Modelos (*Model Evaluate*)

Algumas questões sobre performance de modelos e sua acurácia na escolha das questões podem e devem ser levantadas para poder mensurar a qualidade, eficiência e a eficácia. O Rasa

NLU possui mecanismos para ajudar avaliar o seu modelo (model evaluate).

#### 2.4.2.1 *Intent Classification*

Para compreender melhor algumas questões sobre os modelos, se eles possuem dados suficientes, se as intenções e entidades estão bem estruturadas, o Rasa NLU possui um modo de avaliação (*Evaluation Mode*), que auxilia com estas respostas. O Framework utiliza uma técnica padrão no aprendizado de máquina: manter alguns dados separados como um conjunto de testes (RASA,2018a).

O *script* de classificação das intenções, gera um relatório final com uma matriz de confusão (confusion matrix) e um histograma do modelo avaliado. O relatório registra a precisão (Precision), a chamada (Recall) e a medida F1 (F1-Score) para cada intenção e entidade, além de fornecer uma média geral.

A matriz de confusão mostra quais intenções são confundidas com outras (*True Intent x Predicted Intent*), qualquer amostra que tenha sido predita incorretamente é registrada e salva em um arquivo chamado errors.json para facilitar a depuração.

O histograma permite visualizar a distribuição de confiança para todas as previsões, com o volume de predições corretas e incorretas sendo exibido pelas barras azuis e vermelhas, respectivamente.

#### 2.4.2.2 *Entity Scoring*

Para avaliar a extração de entidades (*evaluate entity extraction*), o algoritmo utiliza uma abordagem baseada em *tags*, considerando apenas as *tags* de tipo entidade por *token*, as *tags* BILOU não são consideradas pelo algoritmo. O ranqueamento acontece a partir da recompensa a extração parcial e não pune a divisão de entidades.

Como exemplo, considere a entidade “perto de Gadelha” e um sistema que extrai a palavra “Gadelha”, o sistema recompensa a extração da palavra “Gadelha” e pune a palavra perdida “perto”, sendo rotulado como uma falha, pois o algoritmo esperaria que a palavra “Gadelha” fosse rotulada como um último token em uma entidade (L-LOC) ao invés de uma única entidade de token (U-LOC). este caso é um exemplo de abordagem baseada no BILOU, a abordagem realizada pelo Rasa NLU, realiza a extração e dividiria a sentença em “perto” e “Gadelha”, dando a cada uma pontuações.



### 3 TRABALHOS RELACIONADOS

Este capítulo apresenta os trabalhos já realizados e com temas em áreas complementares e similares, mostrando a aplicabilidade do uso deste tipo de tecnologia as mais diversas áreas de conhecimento e mercado. Na área educacional, podemos citar os trabalhos de (FERREIRA; UCHÔA,2006), (ANDRADE,2013), (LIMA,2014) e (MORENO et al.,2015), no atendimento ao cliente podemos citar (SOUZA,2018b), (JÚNIOR; BARBOSA,2017), contudo destacamos os trabalho de (FRÖHLICH; SOARES,2018), (MARCOS; MATTOS,2018) e (RODRIGUES, 2018), devido a sua proximida no campo pesquisa.

#### 3.1 Robotização nos relacionamentos: um estudo sobre o uso de *chatbots*

(JÚNIOR; BARBOSA,2017), em seu trabalho, analisam como a robotização de processos está crescendo nas empresas, principalmente nas que possuem departamentos com colaboradores destinados ao relacionamento de clientes. O trabalho explora o quão sensível possa estar o cliente não considerado digital, em uma realidade de atendimento feito por robôs, em um período que cada vez mais, as empresas estão em busca da redução de custos a partir da automação de processos.

Para mensurar a aderência do cliente a esse tipo de atendimento realizado por *chatbots*, os pesquisadores utilizaram uma abordagem quantitativa e qualitativa a partir de questionários, para captar informações e conhecer o ponto de vista dos clientes envolvidos na pesquisa. O questionário foi desenvolvido pelos autores, sendo distribuído de forma digital, seguindo o modelo de dimensões de qualidade em serviços, a tabela3apresenta as dimensões e variáveis analisadas pelos autores.

Dimensão	Variáveis	Autores
Confiabilidade	Ausência de erros durante o atendimento	Fitzsimmons e Fitzsimmons (2014)
Responsividade	Fornecimento imediato do serviço; Efetuar um rápido atendimento.	Lovelock e Wright (2001); Fitzsimmons e Fitzsimmons (2014).
Confiança	Chatbot com o conhecimento necessário sobre o produto/serviço ofertado; Entender as necessidades dos clientes.	Zeithalm, Gremler e Bitner (2014). Kotler e Keller (2012)

**Tabela 3 – Objetivos, variáveis e referências para o método**

A coleta de dados foi realizada com pessoas a partir de 16 anos de idade, utilizando amostra aleatória e não probabilística para validação. Nos dados de amostragem da pesquisa, o perfil de cada público concentra-se da seguinte e forma: 42% com idade de 31 e 40 anos; 21% com idade de 21 e 30 anos; 21% com idade de 41 a 50 anos.

Dentre os resultados do trabalho destacamos que 93% dos respondentes afirmam estar cientes de que em algum momento, ao contatar um serviço de atendimento ao cliente ou semelhante, foram atendidos por um robô. 24% destes atendimentos são realizados a partir de

chats em sites de empresas, 8% dos atendimentos por chats de mídias sociais e 6% em chats por mensagens de texto no celular.

Sobre a qualidade da interação entre o ser humano e os *chatbots*, os autores identificaram que não é completamente interativa, apenas 61% conseguiu interagir satisfatoriamente com o robô. Da amostra que não conseguiu interagir (39%), a dificuldade foi atrelada ao fato do robô não conseguir entender o que o cliente estava falando, representando 63%. Em seguida, 29% demonstrou irritação devido à quantidade de ligações realizadas pelo robô para buscar o contato com eles (10 ligações ou mais), e 8% devido aos horários em que as ligações eram feitas.

Considerando como está a satisfação dos clientes que são atendidos por meio de processos automatizados, após análise e interpretação dos dados teóricos e de campo coletados, que a inteligência agregada a um script de conversação que o robô utilizará para interagir, será determinante para o sucesso do modelo de atendimento a partir da robotização desse processo. As percepções capturadas a partir das respostas reforçam que a linguagem utilizada será o cartão de visitas da empresa. Ou seja, o primeiro contato do cliente em um momento de necessidade para resolver um assunto será a partir deste atendimento.

Como comentam os autores, o mercado está aberto para novas profissões, com a possibilidade de um profissional graduado em Letras ter a chance de entrar para o segmento de tecnologia, atuando como Linguista Computacional. E esse tipo de complementação é factível, desde que o profissional tenha em mente que o mercado inteiro está mudando, e os clientes querem cada vez mais serviços automatizados, soluções ágeis e que tudo seja resolvido com apenas um contato. Existem inúmeras possibilidades e a satisfação dos clientes é o que vai dar a velocidade de toda essa evolução. E mesmo assim, abrem a discussão acerca de como ficam os relacionamentos pessoa-pessoa. Até onde o ser humano está preparado para deixar de estabelecer a conversação entre pessoas, para assumir relacionamentos virtuais?

## **3.2 IBM Watson como Ambiente para Desenvolvimento e Execução de um *Chatbot* – Um Estudo de Caso Aplicado ao Processo de Atendimento ao Usuário**

(MARCOS; MATTOS,2018), busca apresentar um estudo prático da implantação completa de um sistema inteligente, desenvolvendo um Chatbot de atendimento em primeiro nível aos usuários de uma organização. O projeto, teve por objetivo explorar as facilidades técnicas disponíveis. Como resultado, obteve um estudo teórico-prático dos benefícios que este tipo de ambiente possa trazer as corporações e este modelo de negócio.

A metodologia de pesquisa definida pelo autor foi organizada da seguinte forma:

- Estudo e levantamento teórico: definição de métodos e técnicas de implementação a serem

utilizadas;

- Estudo Exploratório: definição de uma plataforma de desenvolvimento para agentes e representação de uma base de conhecimento;
- Estudo de caso: contexto de aplicação e validação dos resultados obtidos;
- Construção do protótipo: validação prática de ambientes de desenvolvimento de *chatbots*;
- Resultados obtidos: validar junto a corporação explorada como estudo de caso.

O Autor, realizou o levantamento dos principais problemas relacionados ao processo de atendimento da empresa, destacou alguns aspectos: desorganização dos atendimentos, falta de histórico, dificuldade de realizar cobranças, acompanhamento, retrabalho e o subutilização de mão de obra técnica. Dentre os principais chamados à equipe, 85% foram categorizados como problemas básicos, que podem ser resolvidos diretamente pelo pessoal do primeiro nível de atendimento, sem qualquer intervenção de pessoal mais técnica.

O Autor identificou que o problema estava no volume excessivo de chamados recorrentes que surgem para equipe de suporte, alinhando a sua proposta deste trabalho, que é utilizar um ambiente corporativo e a problemática como campo de estudo prático para o desenvolvimento de um Chatbot especialista.

Após a etapa de análise de uma plataforma presente no mercado para fornecer a aprendizagem de máquina e processamento em linguagem natural, o autor definiu que o melhor sistema de apoio seria o IBM Watson, pela sua facilidade em gerar conhecimento e tomar decisões com grandes quantidades de dados, incluindo também tipos não estruturados.

O projeto tinha como objetivo o tratamento das informações para alimentar a base e a construção *dobot* propriamente dito. Já que nesta etapa a concentração estava voltada ao fluxo de execução do novo ambiente, foi uma estratégia de projeto, instalação, a execução e os testes do componente construído a partir da própria plataforma Assistant.

A implantação de atendimento por um chatbot, realizou um primeiro nível de atendimento e somente na persistência de um problema o usuário ativaria processos tradicionais, reduzindo a carga de tarefas repetitiva, que segundo estatísticas contribuíram com aproximadamente 78% dos chamados abertos no sistema de atendimento. As próximas etapas, já em construção, consideram novos motores de ativação: por dispositivo móvel e por componente embarcado. E neste mesmo sentido, uma possibilidade de trabalho futuro, é a implantação de mecanismos mais eficientes para tomadas de decisão.

### 3.3 *Enhancing Chatbots using machine learning*

Dissertação de Mestrado de (RODRIGUES,2018), utiliza *Chatbots* no atendimento de *call centers*. Em sua pesquisa, o autor explora a combinação de vários modelos de aprendizagem, com o intuito de encontrar o modelo que melhor se adequa a necessidade da empresa na qual foi realizada a pesquisa. Neste trabalho, o autor explorou a possibilidade de criar um sistema híbrido Humano/Máquina que seja capaz de responder de forma autônoma e automática às perguntas mais frequentes, bem como às perguntas mais exigentes, recorrendo sempre que necessário à ajuda do assistente humano.

Para atingirmos este objetivo, foram testados vários modelos de aprendizagem e vários conjuntos de features para tentarmos encontrar o modelo que melhor se adequa às necessidades das empresas.

Para escolher qual a ferramenta de NLP que seria usada foram submetidos três toolkits para testes: spaCy, NLTK e OpenNLP. Foi construído um avaliador que dado um texto, utilizaria cada um dos toolkits para gerar uma solução a ser rankeada. Após o teste, foram criadas matrizes de confusão para cada um dos toolkits, bem como os totais de erro para cada uma das categorias, sendo o spaCy a ferramenta escolhida para o trabalho, partindo para a definição dos modelos.

Modelo 01, treinado e implementado para trabalhar com a informação do dataset do SemEval. teve comportamento errático onde pequenas alterações na pergunta resultavam em altas variabilidades nos resultados não deixavam confortáveis quem tinha de ir para a rua vender o produto e fazer demonstrações com ele.

O modelo 02, compara os vetores de duas frases quando ambas as frases têm mais de um elemento na categoria, utilizando semelhança de comparação (Cosine Similarity) em questão é claramente mais conservador e apresenta resultados mais baixos durante o teste. Mas o facto de ser mais previsível, permitia a quem vendia os produtos e tinha um guião do que mostrar nas apresentações, que ficasse mais descansado e confiante com os resultados.

O modelo 03 modelo genérico de representação semântica foi treinado com o dataset do SNLI para a tarefa de semelhança semântica. Utiliza bi-directional LSTM com recurso a caracterização das palavras em vetores feito pelo gloVe. Esteve implementado durante um curto período de tempo porque a partir da criação deste, foi desenvolvido o modelo 04, que teve melhores resultados em toda a linha.

O modelo 04 foi um modelo final, utilizando o cosine similarity, sendo uma versão melhorada do modelo 03 somada a técnica utilizada no modelo 02. Conclui-se aqui que os resultados obtidos no Modelo 04 foram melhores em toda a linha, e comparação aos modelos anteriores.

O autor conclui que a utilização da similaridade semântica para a construção da knowledgebase como uma boa técnica. A existência de uma grande receptividade da parte dos operadores

às várias iterações das provas de conceito realizadas, mostram a atratividade da tecnologia e das funcionalidades para o mercado. Como resultados de retenção, houve uma redução em 80% numa fase embrionária do sistema indica que é possível reduzir o custo do atendimento ao cliente.

## 4 PROJETO

### 4.1 Coleta de Dados

Para o levantamento das ocorrências de atendimento, foram coletados dados de atendimento da empresa, com as equipes de vendas (marketing) e atendimento. Ambas as equipes, até o momento da escrita do presente trabalho, estão sediadas na cidade de Manaus, estado do Amazonas.

A cada atendimento realizado no setor, assim que concluído o atendente em uma planilha compartilhada entre os atendentes do setor, inseria um resumo do atendimento, os dados do cliente atendido, a data de ocorrência, o canal de atendimento e se o problema do cliente solucionado.

Foram coletados dados das ocorrências de atendimento referentes aos meses de abril, maio e junho de 2018, não sendo possível coletar dados anteriores por não haver um processo de coleta de dados de atendimento. Após coletados, os atendimentos foram agrupados de acordo com o canal (telefone, email, sms, facebook e whatsapp) e foram dispostos em planilhas específicas, nas quais cada linha representava uma ocorrência de atendimento, com informações sobre: o cliente, o motivo do atendimento, a data da ocorrência e quaisquer observações fornecidas pelo atendente.

Ao final, foram selecionados 357 ocorrências de atendimentos, percentences ao canal *Facebook Messenger*, este canal foi escolhido por já ser um canal de comunicação ativo da empresa, por permitir a automação de atendimento por chatbots e ter menor impacto direto no setor: por ser considerado um canal secundário da empresa, ficando atrás apenas de telefone e whatsapp. Finalizados os agrupamentos, os dados foram novamente classificados de acordo com os principais ocorrências de atendimento para aplicação do princípio de pareto e construção do diagrama.

### 4.2 Classificação e Construção do Diagrama de Pareto

O Princípio de Pareto, também conhecido como lei dos 80/20, diz que um pequeno número de causas (aproximadamente 20%) é responsável pela maioria dos problemas (aproximadamente 80%). Esse princípio serve de base para o Diagrama de Pareto. O diagrama de Pareto é uma forma de descrição gráfica onde busca-se identificar quais itens são responsáveis pela maior parcela dos problemas (NEUMANN,2012).

O Diagrama é uma ferramenta que permite fácil visualização e identificação das causas ou problemas mais recorrentes, possibilitando maior concentração de esforços sobre os mesmos (NEUMANN,2012). A construção do diagrama aconteceu em duas etapas: na primeira etapa, as ocorrências são agrupadas e categorizadas em suas respectivas categorias e ao final, contabiliza-

das. Na segunda etapa, são calculados os percentuais de cada evento em relação ao somatório dos eventos ocorridos (soma do total de ocorrências) e o percentual acumulado, que nada mais é que a soma dos percentuais individuais de cada categoria. A Tabela4 apresenta às 357 ocorrências já categorizadas, com percentual de cada grupo de ocorrência e percentual acumulado.

CATEGORIA	TOTAL	PERCENTUAL	ACUMULADO
NOVO REVENDEDOR	87	24,3697479	24,3697479
SAÍDA DE MERCADORIA	60	16,80672269	41,17647059
ACORDO DE PAGAMENTO	42	11,76470588	52,94117647
AVISOS E BOLETOS	35	9,803921569	62,74509804
INFORMAÇÃO	26	7,282913165	70,0280112
CONFIRMAÇÃO DE PAGAMENTO	21	5,882352941	75,91036415
INFORMAÇÃO/DÚVIDA	14	3,921568627	79,83193277
CATÁLOGO EM PDF	13	3,641456583	83,47338936
INDENIZAÇÃO	10	2,801120448	86,2745098
CADASTRO WEB	10	2,801120448	89,07563025
RECLAMAÇÃO	9	2,521008403	91,59663866
CHEGADA DE MERCADORIA	8	2,240896359	93,83753501
ALTERAÇÃO DE CADASTRO	6	1,680672269	95,51820728
CONFIRMAÇÃO DE ENVIO DE PEDIDO	5	1,400560224	96,91876751
ENVIO CATALOGO PDF	3	0,840336134	97,75910364
INTERESSE EM SER PROMOTOR	3	0,840336134	98,59943978
PAGAMENTO CARTÃO DE CRÉDITO	2	0,56022409	99,15966387
INFORMAÇÃO/PEDIDOS	2	0,56022409	99,71988796
ERRO DE PAGAMENTO	1	0,280112045	100

**Tabela 4 – Categorização dos atendimentos pelo canal Facebook Messenger**

A primeira coluna da Tabela4, apresenta a categoria da ocorrência; a segunda coluna, o total de ocorrências nesta categoria; a terceira coluna, apresenta o percentual de ocorrências da categoria em relação ao total de ocorrências catalogadas; e na última coluna, o acumulado do percentual de ocorrências.

Para a construção do diagrama de Pareto, foram usadas as bibliotecas *Pandas*<sup>1</sup> e *NumPy*<sup>2</sup> na importação, limpeza e classificação dos dados e plotados utilizando a biblioteca *matplotlib*<sup>3</sup>, o *script* de análise foi construído em linguagem *Python*<sup>4</sup> e diagrama obtido pode ser visualizado no gráfico2.

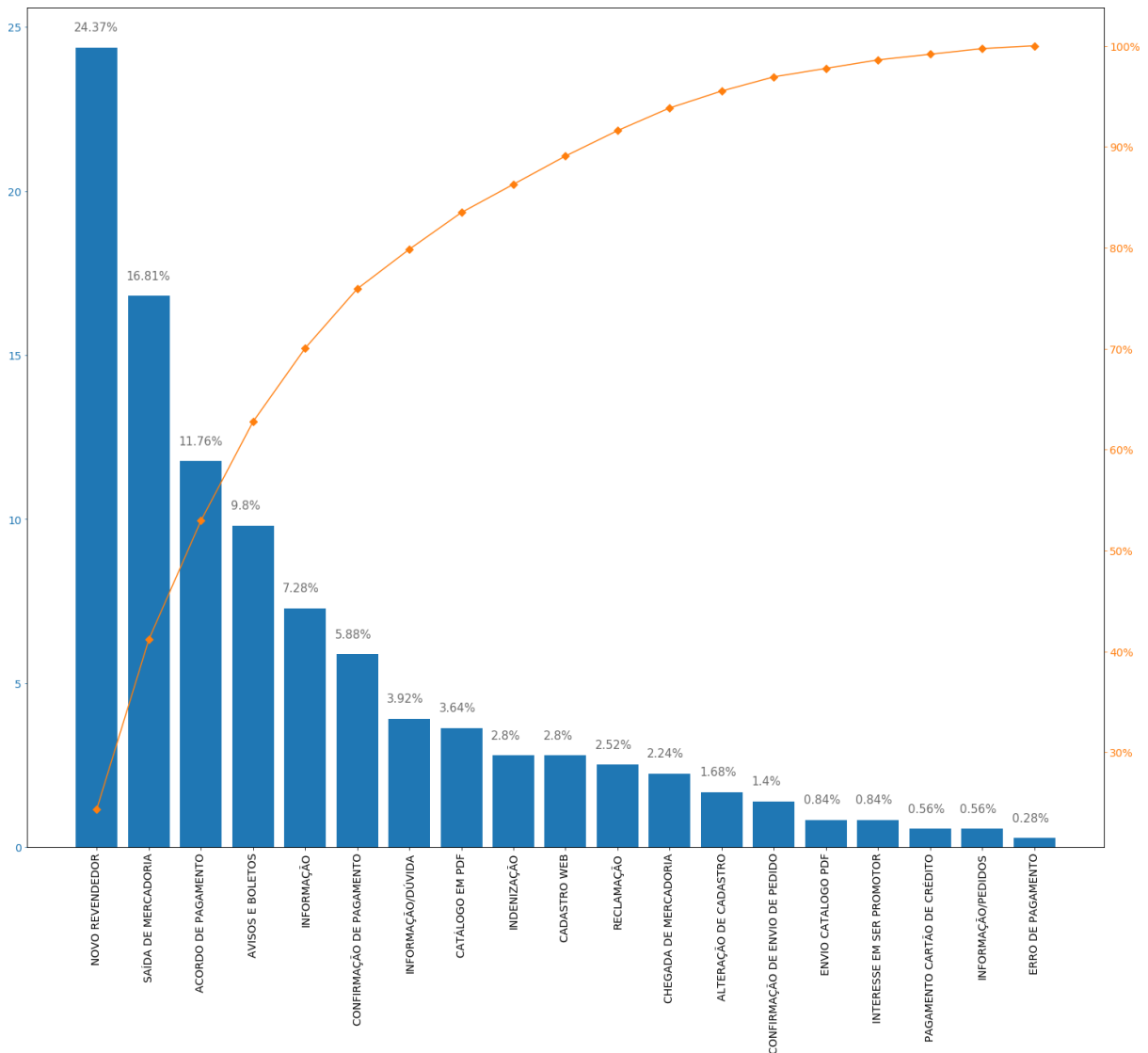
O gráfico2, apresenta o diagrama de pareto completo, que nada mais é que a representação gráfica do método de pareto aplicado às ocorrências de atendimento. É possível inferir após leitura e análise do diagrama que as ocorrências sobre Novo Revendedor, Saída de Mercadoria e Acordos de pagamento são as categorias com maior ocorrências.

<sup>1</sup> <https://pandas.pydata.org>

<sup>2</sup> <http://numpy.org>

<sup>3</sup> <http://matplotlib.org/>

<sup>4</sup> <https://www.python.org>



**Gráfico 2 – Diagrama de Pareto**

O *chatbot* irá responder inicialmente às questões relacionadas à Saída de Mercadoria, Novo Revendedor e Avisos e boletos, prioritariamente na categoria Novo Revendedor por ser a categoria com maior percentual acumulado, com pouco mais de 24% e conseqüentemente, com maior impacto no atendimento. Consideramos que Acordo de Pagamentos, apesar do grande volume ainda deve ser realizado por um humano e questões sobre esta categoria devem ser direcionadas à equipe de atendimento.

### 4.3 Definição dos Modelos

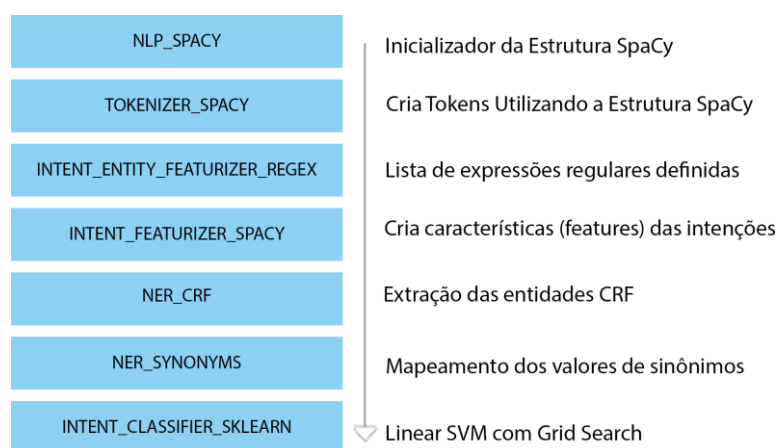
O Rasa NLU, possui uma estrutura componentizada e oferece várias formas de combinação de seus componentes, sendo facilmente adaptado aos mais diversos contextos. Dentre as possibilidades de modelos a serem construídos, 4 configurações de **pipeline**<sup>5</sup> foram com-

<sup>5</sup> [https://rasa.com/docs/nlu/choosing\\_pipeline/](https://rasa.com/docs/nlu/choosing_pipeline/)



postas, com intuito de selecionar o modelo com melhor desempenho e adequação ao projeto. Dentre os modelos propostos, dois modelos utilizam vetores pré-treinados: com `nlp_spacy`<sup>6</sup> + `intent_classifier_sklearn`<sup>7</sup> e utilizando `nlp_mitie`<sup>8</sup> + `intent_classifier_sklearn`, respectivamente; os outros dois modelos não utilizam vetores pré-treinados, utilizando o componente `tensorflow_embedding`, variando apenas a sua configuração de uso, ficando organizados da seguinte forma:

O **modelo 01** utiliza os componentes do `Spacy`<sup>9</sup> e os classificadores do `Scikit-Learn`<sup>10</sup>, o Modelo de palavras utilizado foi o "pt\_core\_news\_sm" com vocabulários, sintaxes e entidades pré-treinadas. A figura 5, apresenta a arquitetura de fluxo dos componentes e de maneira simplificada, sua responsabilidade dentro do conjunto proposto.



**Figura 5 – Arquitetura - Modelo 01**

#### Pipeline - Modelo 01

```

pipeline : 1
- name : "nlp_spacy" 2
- name : "tokenizer_spacy" 3
- name : "intent_entity_featurizer_regex" 4
- name : "intent_featurizer_spacy" 5
- name : "ner_crf" 6
- name : "ner_synonyms" 7
- name : "intent_classifier_sklearn" 8

```

<sup>6</sup> <https://rasa.com/docs/nlu/components/#nlp-spacy>

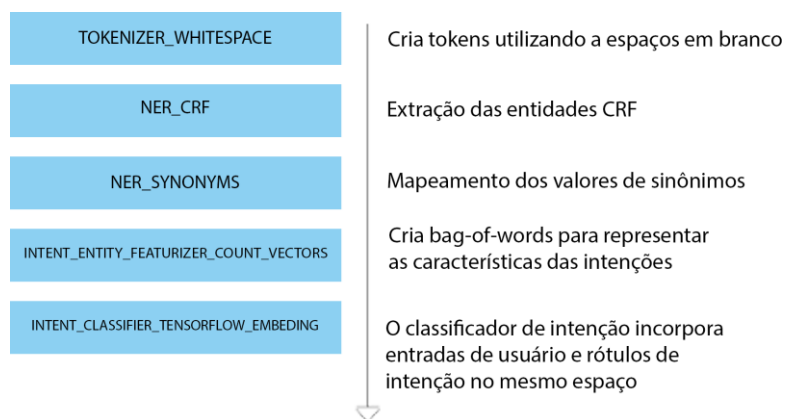
<sup>7</sup> <https://rasa.com/docs/nlu/components/#intent-classifier-sklearn>

<sup>8</sup> <https://rasa.com/docs/nlu/components/#nlp-mitie>

<sup>9</sup> <http://spacy.io>

<sup>10</sup> <http://scikit-learn.org>

O modelo 02 utiliza o componente embutido do tensorflow <sup>11</sup> na classificação de intenções associado aos componentes do Spacy para tokenização, como um tokenizador por espaços em branco, NER (Named Entity Recognition) para para os Random Fields e NER para sinônimos. A figura6, apresenta a arquitetura de fluxo dos componentes e sua responsabilidade dentro do conjunto proposto.



**Figura 6 – Arquitetura - Modelo 02**

**Pipeline - Modelo 02**

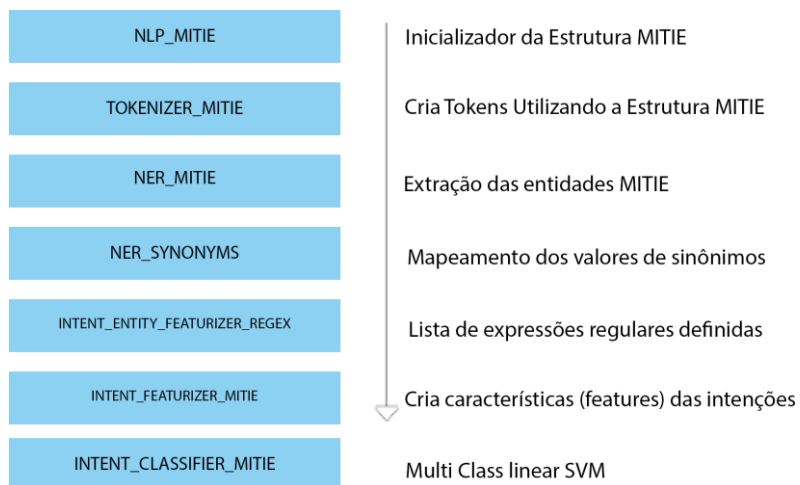
```

pipeline :
- name : " tokenizer_whitespace "
- name : " ner_crf "
- name : " ner_synonyms "
- name : " intent_featurizer_count_vectors "
- name : " intent_classifier_tensorflow_embedding "

```

O modelo 03 utiliza a biblioteca e componentes MITIE<sup>12</sup>, desenvolvida em dlib <sup>13</sup> é considerada uma biblioteca de alta performance em machine learning, oferecendo ferramentas de Named Entity Recognition (NER) e Binary Relation Detection (BED). O uso da biblioteca exige que você tenha um Corpus de palavras, o que é muito interessante no caso de línguas não disponíveis no spacy, permitindo uma construção de vocabulário totalmente personalizada. O corpus utilizado foi desenvolvido por Fabio Mazzo <sup>14</sup>, a partir de artigos da Wikipedia e ao Corpus Laps da UFPA <sup>15</sup>. A figura7, apresenta de forma simplificada a arquitetura de fluxo dos componentes dentro do modelo proposto.

<sup>11</sup> <https://www.tensorflow.org>  
<sup>12</sup> <https://github.com/mit-nlp/MITIE>  
<sup>13</sup> <http://dlib.net>  
<sup>14</sup> <https://github.com/fabiocmazzo/corpusportugues>  
<sup>15</sup> Universidade Federal do Pará



**Figura 7 – Arquitetura - Modelo 03**

#### Pipeline - Modelo 03

```

pipeline : 1
- name: "nlp_mitie" 2
  model: "data/total_word_feature_extractor.dat" 3
- name: "tokenizer_mitie" 4
- name: "ner_mitie" 5
- name: "ner_synonyms" 6
- name: "intent_entity_featurizer_regex" 7
- name: "intent_featurizer_mitie" 8
- name: "intent_classifier_sklearn" 9

```

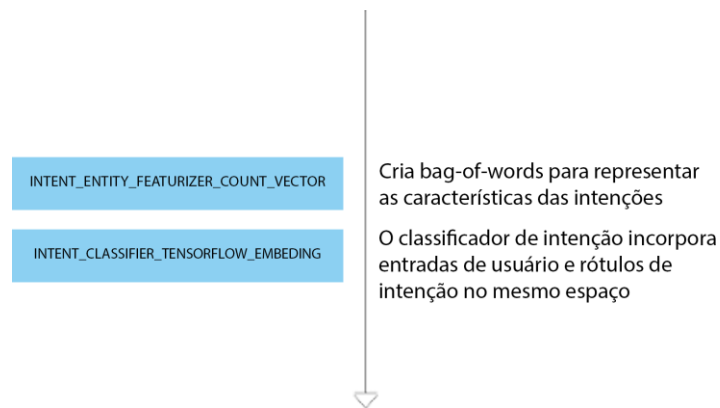
O modelo 04 também utiliza uma abordagem com tensorflow, mas apenas com um componente do sklearn. Esta abordagem é menos genérica e mais específica no uso dos componentes, utilizando o analisador char\_wb do sklearn e o classificador de intenções embutido do tensorflow. A figura5, apresenta a arquitetura de fluxo dos componentes e de maneira simplificada, sua responsabilidade dentro do conjunto proposto.

#### Pipeline - Modelo 04

```

pipeline : 1
- name: "intent_featurizer_count_vectors" 2
  analyzer: char_wb 3
  min_df: 0.006789048157425257 4
  max_df: 0.4343982144945721 5
  max_ngram: 7 6
- name: "intent_classifier_tensorflow_embedding" 7
  epochs: 34 8
  batch_size: 165 9

```



**Figura 8 – Arquitetura - Modelo 04**

embed_dim :	60	10
C2:	0.0002167140015537944	11
C_emb:	0.00022451701750527038	12
droprate :	0.17386092794138916	13
num_hidden_layers_a :	0	14
hidden_layer_size_a :	170	15
num_hidden_layers_b :	4	16
hidden_layer_size_b :	130	17

#### 4.3.1 Critérios de Seleção

Definidos os modelos a serem utilizados no projeto, com intuito de selecionar o melhor modelo a ser utilizado, foram adotados os seguintes critérios de seleção no processo de seleção do modelo:

- Flexibilização;
- Acurácia;
- Assertividade;
- Conseguir trabalhar com um Dataset inicial pequeno.

Para todos os modelos, foram utilizadas as mesmas políticas (*policies*): 300 épocas, máximo de 5 histórias, confiança (*confidence*) mínima das intenções e diálogos preditos de 0.3 e memorização máxima de 5 histórias. Para a análise dos resultados obtidos, foram geradas

matrizes de confusão (*confusion matrix*) e histogramas da distribuição de confiança na predição de intenções do modelo.

Para validação dos modelos, a biblioteca sklearn (Scikit-Learn) foi utilizada para a avaliação (*evaluation*) e como métricas, foram adotados o *F1-Score*, *Precision* e *Accuracy* <sup>16</sup>, todas estas métricas foram adotadas tanto para testes quanto para treino dos modelos, sendo a pontuação de teste a de maior relevância dentre as métricas de avaliação.

A métrica de precisão (*precision*) é a relação de **tp** onde **tp** é o número de positivos verdadeiros (*true positives*) e **fp** o número de falsos positivos (*false positives*). A precisão é a capacidade do classificador de não rotular como positiva uma amostra que é negativa. A métrica pode ser definida pela equação:

$$precision = tp / (tp + fp)$$

A métrica de chamada *Recall* é a relação **tp** / (**tp** + **fn**) onde **tp** é o número de positivos verdadeiros (*true positives*) e **fn** o número de falsos negativos (*false negatives*). O *Recall* é intuitivamente a habilidade do classificador de encontrar todas as amostras positivas. A métrica pode ser representada pela equação:

$$recall = tp / (tp + fn)$$

A Pontuação F1 ou *F1-Score* é a média ponderada da precisão (*precision*) e chamada (*recall*). Essa pontuação leva em conta tanto os falsos positivos (fp) quanto os falsos negativos (fn). O *F1-Score* pode ser definido pela equação:

$$F1 = 2 * (Recall * Precision) / (Recall + Precision)$$

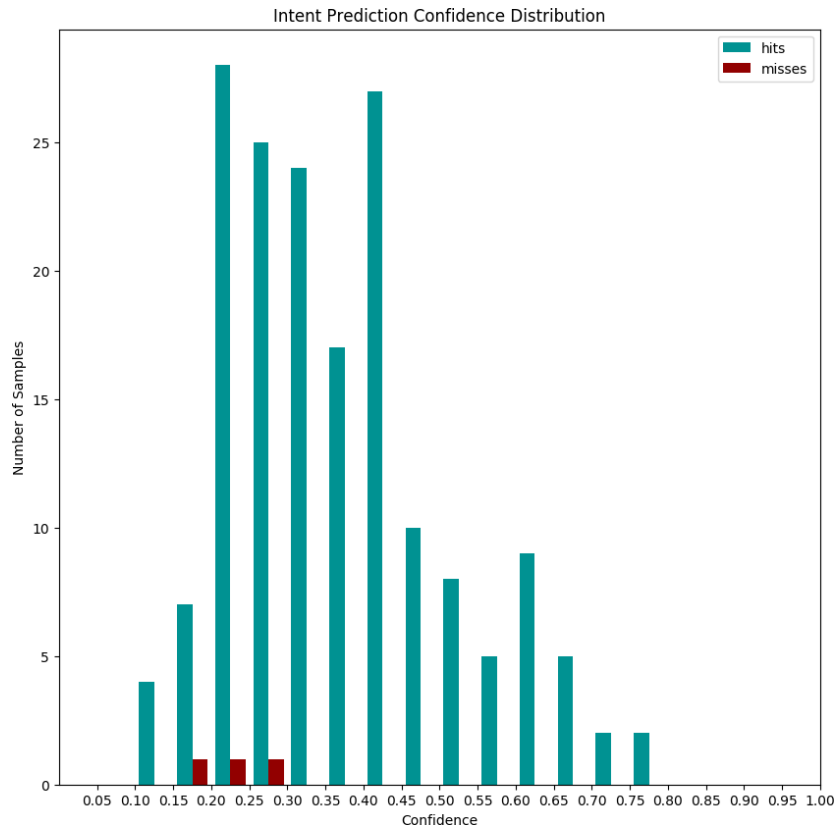
### 4.3.2 Testes e Validação dos Modelos

Para validação dos modelo de configuração de componentes, cada modelo foi treinado utilizando o mesmo dataset construído a partir dos atendimentos realizados. O Dataset continha 371 expressões relacionadas as atividades do negócio, com intenções, entidades e sinônimos já classificados e listados no arquivo de configuração comum à todos.

#### 4.3.2.1 Modelo 01

Como pode ser analisado no histograma disponível na Figura3e na Matriz de Confusão na figura9, o modelo obteve o maior número de predições entre 0.20 e 0.45 de confiança, além de errar algumas predições com uma acurácia abaixo de 0.3, que é o limite máximo de resposta permitido.

<sup>16</sup> [https://scikit-learn.org/stable/modules/generated/sklearn.metrics.precision\\_recall\\_fscore\\_support.html](https://scikit-learn.org/stable/modules/generated/sklearn.metrics.precision_recall_fscore_support.html)



**Gráfico 3 – Histograma Pipeline 01**

Intent Evaluation Results					
train Accuracy:	train F1-score	train Precision	test Accuracy	test F1-score	test Precision
1.000 (0.000)	1.000 (0.000)	1.000 (0.000)	0.650 (0.104)	0.634 (0.102)	0.646 (0.102)

**Tabela 5 – Intent Evaluation Results - Modelo 01**

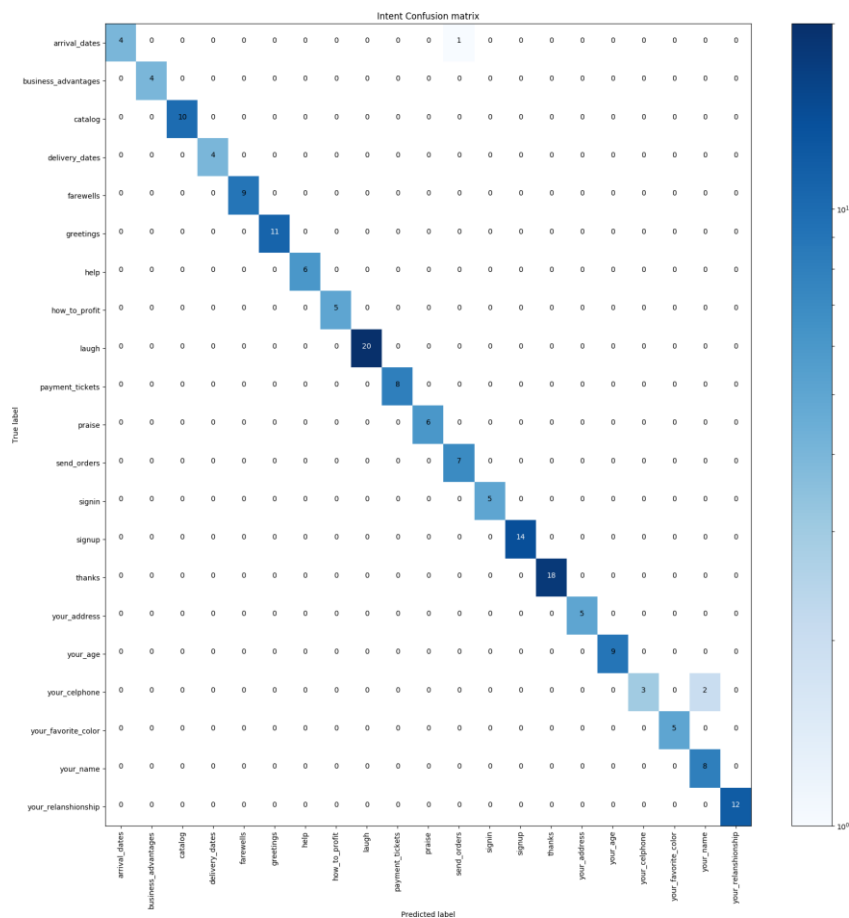
A matrix de confusão presente na Figura9, dispõe os dados preditos de forma visual e de análise mais amigável, no eixo x temos as intenções preditas e no eixo y as intenções reais (verdadeiras). É possível verificar que o modelo errou a predição da data de chegada do pedidos, e o número de contato, quando questionado.

A tabela5, apresenta as métricas de Acurácia (Accuracy), F1-Score e Precisão (Precision), para treino e para testes do modelo. Os resultados de treino do Modelo 01 foram excelentes, obtendo a pontuação máxima, contudo, nos testes, a pontuação média foi relativamente inferior comparadas as de treino, com uma média de 0.643.

#### 4.3.2.2 Modelo 02

O histograma disponível na Figura5, identificamos uma significativa melhora nos resultados do histograma, a confiança na predição está entre 0.9 e 1.00, acima de 120 amostras de testes realizados, dando impressões positivas sobre o modelo.

Seguido para a análise da matriz de confusão disponível na figura10, confirmamos que o



**Figura 9 – Matriz de confusão Pipeline 01**

Intent Evaluation Results					
train Accuracy	train F1-score	train Precision	test Accuracy	test F1-score	test Precision
0.958 (0.020)	0.949 (0.028)	0.949 (0.032)	0.699 (0.076)	0.958 (0.020)	0.949 (0.028)

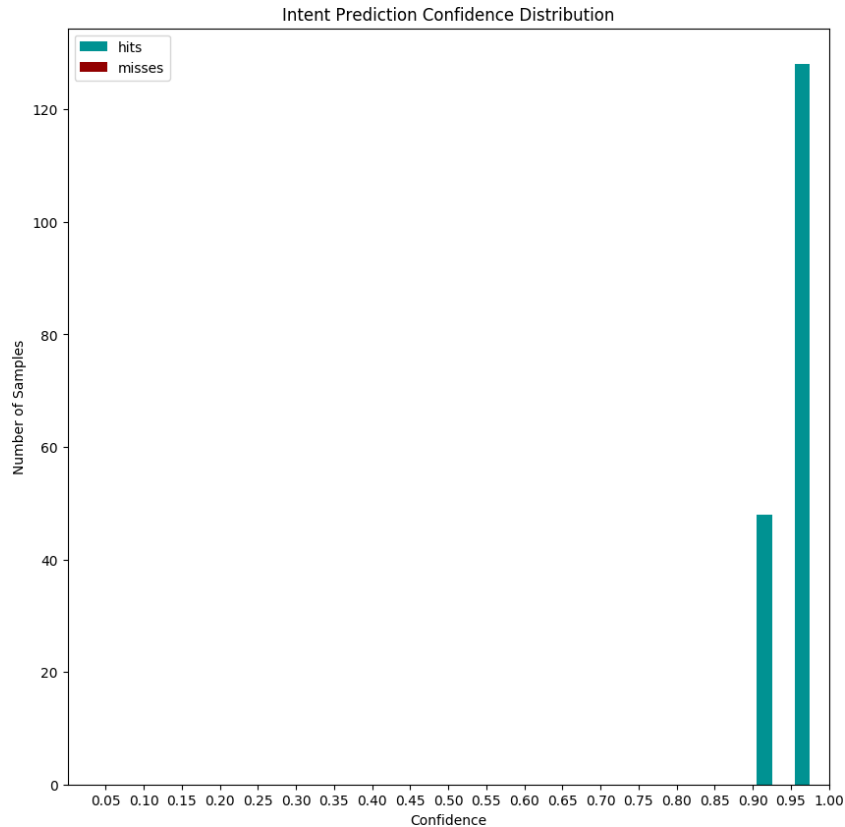
**Tabela 6 – Intent Evaluation Results - Pipeline 02**

modelo não errou nenhuma predição nos testes com um Dataset pequeno, nos levando a concluir com bom resultados de análise e partindo para a análise das métricas.

Finalizamos com as métricas de treino e testes, disponível na tabela6, o modelo no treino não teve uma acurácia tão boa quanto no modelo anterior, mas ainda assim é uma boa acurácia, tendo a mínima em 9.49. Nos testes, ele apresentou uma variância, tendo uma acurácia de 0.699 e F1-Score de 0.958 e Precisão de 0.949, consideramos que o modelo teve bons resultados em relação ao modelo anterior.

#### 4.3.2.3 Modelo 03

O histograma gerado a partir do Modelo 03, disponível na Figura5, apresentou uma grande dispersão na confiança de respostas, mas apesar de ser um modelo com dicionário de palavras disponível, teve a maior concentração de confiança das respostas em 0.25, ficando abaixo do ideal proposto que foi de 0.30.



**Gráfico 4 – Histograma Pipeline 02**

Intent Evaluation Results					
train Accuracy	train F1-score	train Precision	test Accuracy	test F1-score	test Precision
1.000 (0.000)	1.000 (0.000)	1.000 (0.000)	0.661 (0.080)	0.644 (0.087)	0.656 (0.104)

**Tabela 7 – Intent Evaluation Results - Pipeline 03**

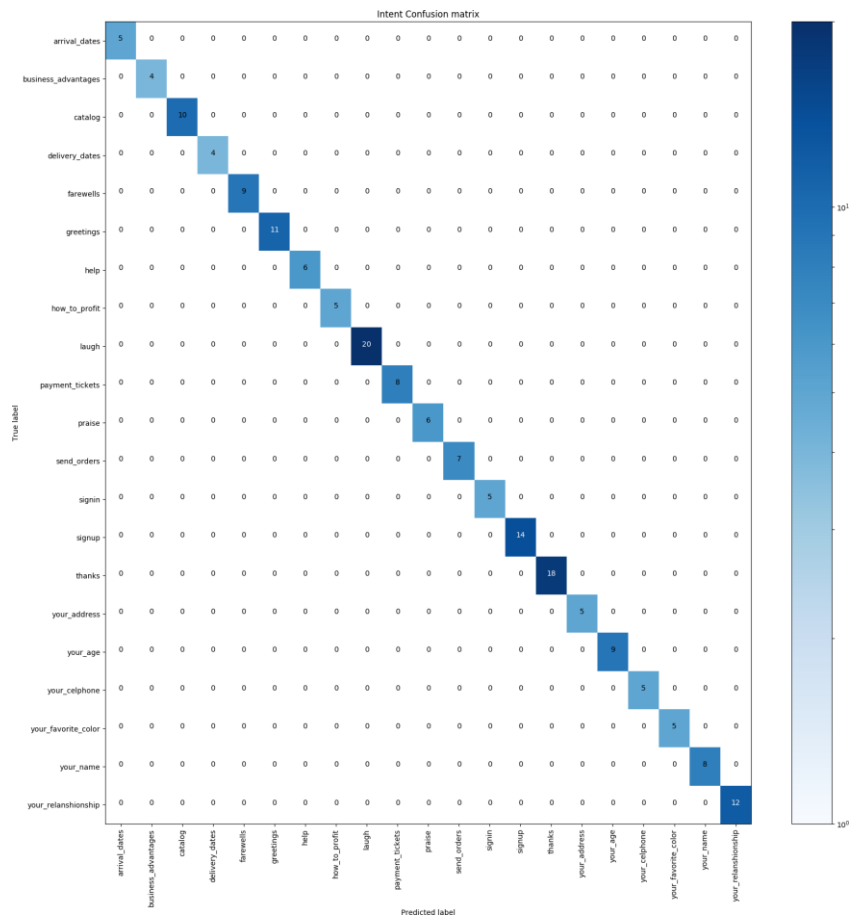
A partir da Matriz de confusão do modelo, disponível na figura11, podemos afirmar que apesar da baixa confiança nas predições, o Modelo 03, não errou nenhuma das predições feitas por ele no testes. Podemos verificar pelo mapa de calos que apenas para algumas intenções acerto foi relativamente alto, como agradecimento e cadastro.

A análise com as métricas disponíveis na Tabela7, as métricas de treino foram excelentes, mostrando a estabilidade da ferramenta no aprendizado, contudo, nos teste, houve uma significativa variação, mas relativamente próximo aos dados obtidos nos modelos anteriores, tendo nota inferior apenas ao Modelo 02, até então.

#### 4.3.2.4 Modelo 04

Dando prosseguimento ao processo de análise, a partir do histograma do Modelo 04, disponível na Figura6, é um dos modelos com confiança nas respostas mais altas, tal como o modelo 02, com respostas confiança mínima pouco acima de 0.90. Diferente do modelo 02, o modelo 04 possui uma maior concentração em 0.90 e o modelo 02 em 0.95, dando uma precisão





**Figura 10 – Matriz de confusão Pipeline 02**

Intent Evaluation Results					
train Accuracy	train F1-score	train Precision	test Accuracy	test F1-score	test Precision
0.997 (0.004)	0.997 (0.004)	0.997 (0.004)	0.822 (0.061)	0.781 (0.071)	0.762 (0.080)

**Tabela 8 – Intent Evaluation Results - Pipeline 04**

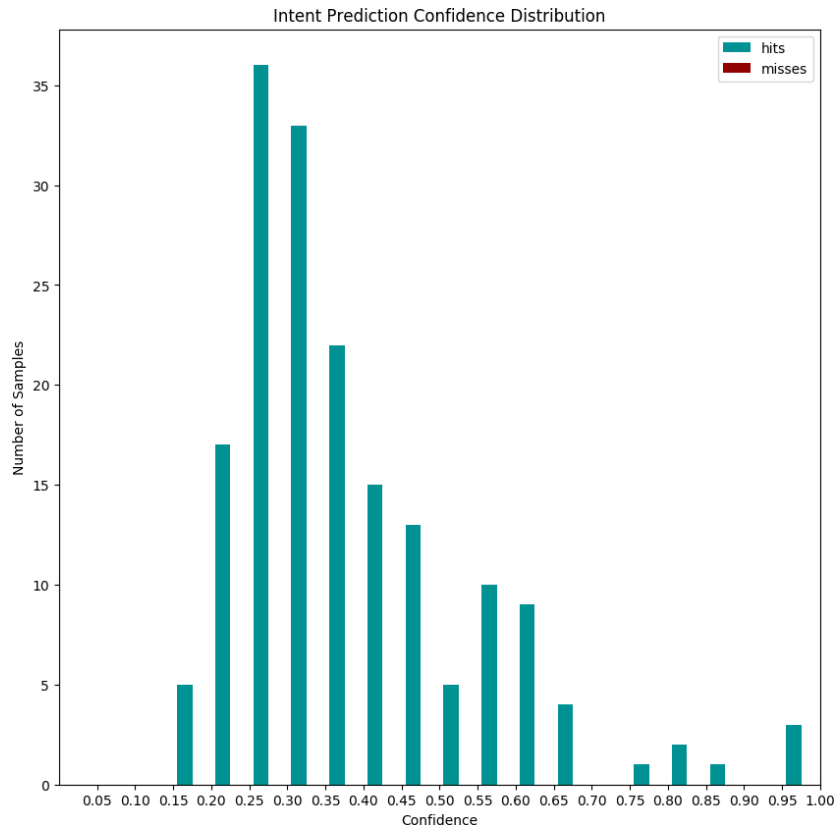
ligeiramente inferior.

Prosseguindo para a matriz de confusão disponível na figura12, podemos afirmar que o modelo não errou nenhuma predição feita, mostrando bons resultados até então, damos continuidade e seguimos para a tabela com as métricas de treino e teste.

Tal com o modelo 02, o modelo 04 não obteve a nota máxima nos treinos do modelo, com uma média de 0.997 para todas as métricas. Nos teste, o modelo, em acurácia mostrou-se o mais efetivo, tendo desempenho superior aos modelos 01 e 03, nos demais aspectos, ficando apenas atrás do modelo 02 no F1-Score e Precision.

#### 4.3.2.5 Resultado dos Testes de Validação

Após realização do processo de análise dos histogramas e Matrizes de Confusão de cada modelo e explorar as métricas treino e testes de cada um, conseguimos decidir a partir destas



**Gráfico 5 – Histograma Pipeline 03**

análises e chegamos aos valores respectivo de cada métrica anteriormente definida.

Concluimos a partir das análises dos histogramas e Matrizes de confusões que os Modelos 02, 03 04, apesar dos índices de confiança diferentes, não erraram nenhuma previsão do modelo, mostrando-se eficazes. Contudo, os modelos 02 e 04, obtiveram maiores índices de confiança dentre os modelos analisados, tendo confiança mínima de 0.90, sendo que o modelo 02 teve uma maior concentração em 0.95.

A tabela9dispõe os resultados de acordo com o modelo testado, com cada métrica analisada discriminada. Concluindo a análise dos modelos, considerando como peso maior a métrica de testes, selecionamos o **Modelo 02**, por apresentar melhor desempenho dentre todos os modelos testados. Partimos para a etapa seguinte, para implantação do modelo e testes com os usuários.

	Modelo 01	Modelo 02	Modelo 03	Modelo 04
train Accuracy	1.000 (0.000)	0.958 (0.020)	1.000 (0.000)	0.997 (0.004)
train F1-score	1.000 (0.000)	0.949 (0.028)	1.000 (0.000)	0.997 (0.004)
train Precision	1.000 (0.000)	0.949 (0.032)	1.000 (0.000)	0.997 (0.004)
test Accuracy	0.650 (0.104)	0.699 (0.076)	0.661 (0.080)	0.822 (0.061)
test F1-score	0.634 (0.102)	0.958 (0.020)	0.644 (0.087)	0.781 (0.071)
test Precision	0.646 (0.102)	0.949 (0.028)	0.656 (0.104)	0.762 (0.080)

**Tabela 9 – Análise dos resultados dos classificadores de intenções**

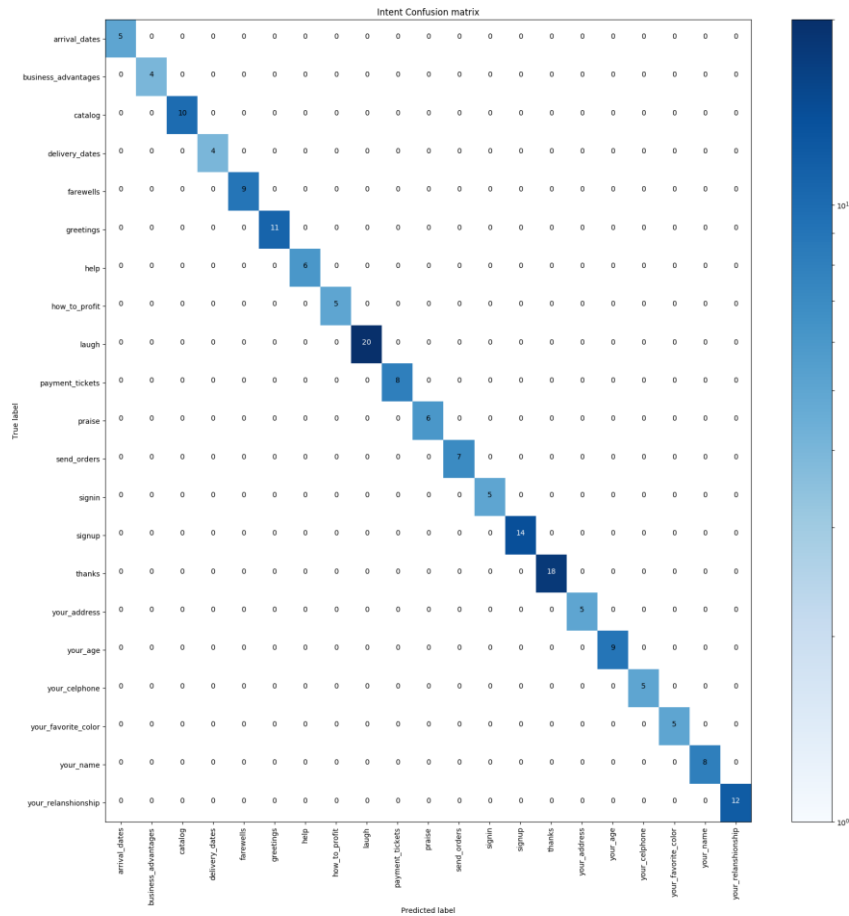


Figura 11 – Matriz de confusão Pipeline 03

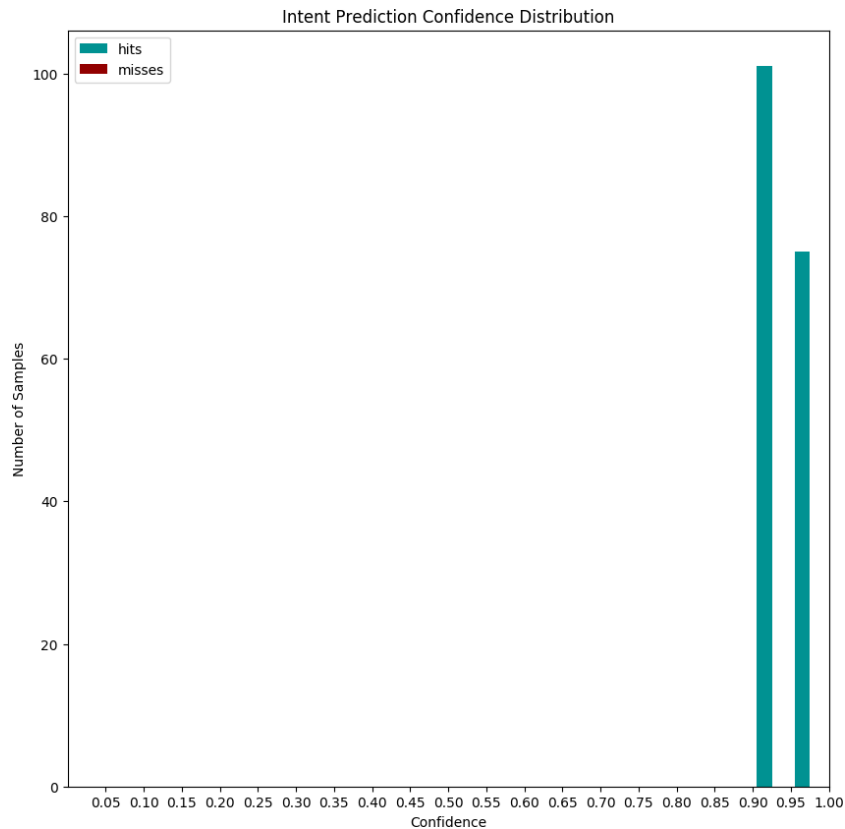
#### 4.3.3 Definições do agente

Projetar experiência de interface de usuário conversacional é complicado porque a conversa vem com muitas expectativas. Quando essas expectativas são satisfeitas, sentimos que a interface é natural, mas, uma vez violadas, sentimos que algo está errado. A última década testemunhou tecnologias e comportamentos da linguagem humana (FADHIL,2018). Segundo (FADHIL,2018) e (DAMIÃO et al.,2014), alguns princípios são importantes do *design* de agentes, dentre os aspectos elencados pelo autor, alguns foram destacados e implantados no *chatbot*:

*Personalidade*: a personalidade do *chatbot* é o que torna uma conversa natural com um indivíduo e transmite maior conforto na conversa. Construir uma personalidade apropriada pode melhorar a experiência do usuário e o seu engajamento.

*Flexibilidade na resposta*: deve fornecer valores flexíveis para várias solicitações do usuário. Por exemplo, fornecer mensagens de erro diferentes como resposta à recorrência de erros similares.

*Simplicidade na interação*: Além do aspecto aparentemente humano da comunicação no tom certo e da previsão e aprendizagem do usuário por meio de uma tarefa.



**Gráfico 6 – Histograma Pipeline 04**

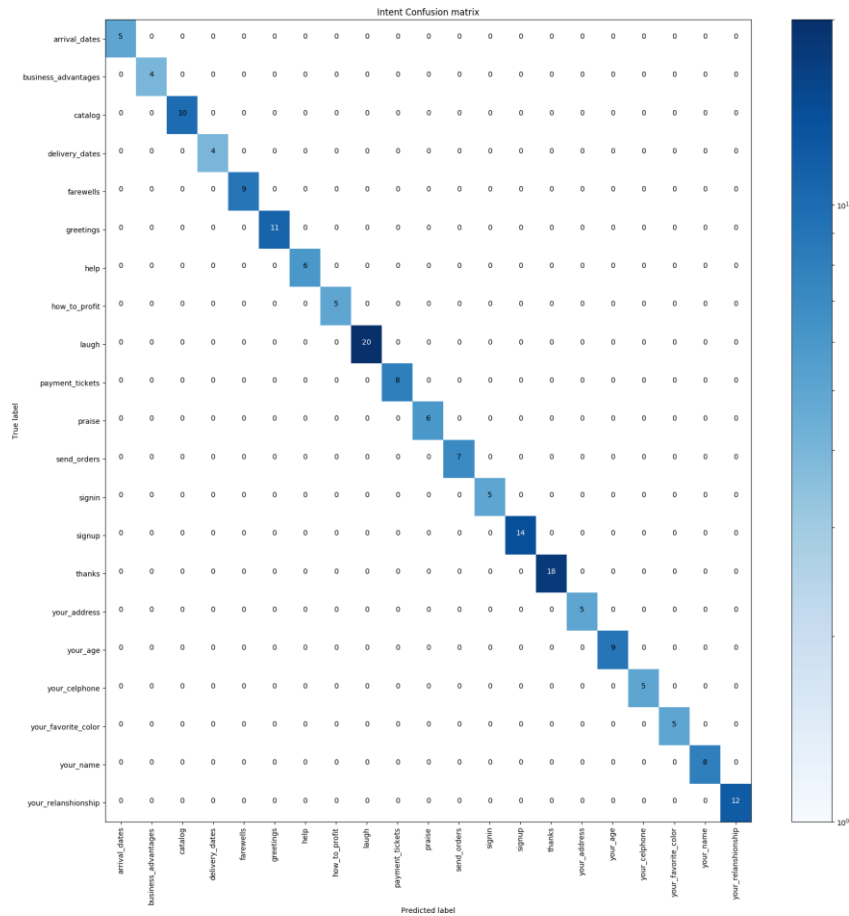
*Tarefas e especificações de serviço:* os *chatbots* devem atender a tarefas específicas e claramente projetadas.

*Texto e botões personalizados:* os usuários não devem ser colocados em uma situação em que tenham que adivinhar as informações corretas necessárias para prosseguir. Além disso, o *bot* deve suportar um dicionário com base em sinônimos para obter o mesmo resultado para vocabulários, como "compra" e "comprar" ou "clientes" e "cliente".

*Fornecer uma saída:* Os seres humanos cometem erros, portanto, os usuários de *bot* devem sempre ser capazes de recomeçar, fazer alterações ou escapar completamente quando ocorre um erro.

A partir destas diretrizes, os seguintes aspectos de interação foram modelados: A personalidade do agente foi definida como uma mulher chamada Lu, com aproximadamente 25 anos. O intuito do agente é ser um guia das dos clientes, com um tom afetuoso, objetivo e paciente. Empenhada em atender com excelência e qualidade no objetivo de solucionar as dificuldades das revendedoras.

As tarefas a serem realizadas pela Lu, são relacionadas aos problemas identificados a partir do diagrama de Pareto, prioritariamente, as de auxílio à Nova Revendedora. O processo de interação, para o fluxo de conversação, foi fundamentado na modelagem do atendimento realizado pela equipe, assim como a linguagem foi aproximada para a linguagem do agente



**Figura 12 – Matriz de confusão Pipeline 04**

proposto. Algumas intenções, foram definidas com múltiplas respostas que levam, mas sempre direcionadas a mesma intenção. Como exemplo, a intenção de despedida, ela pode emitir respostas como "adeus", "até mais", "tchau, tchau".

Em caso de erros de identificação da intenção, como válvula de escapes é sempre possível reiniciar a conversa e voltar ao fluxo inicial e em caso de erros contínuos, a conversa pode ser transferida para uma pessoa do setor de atendimento.

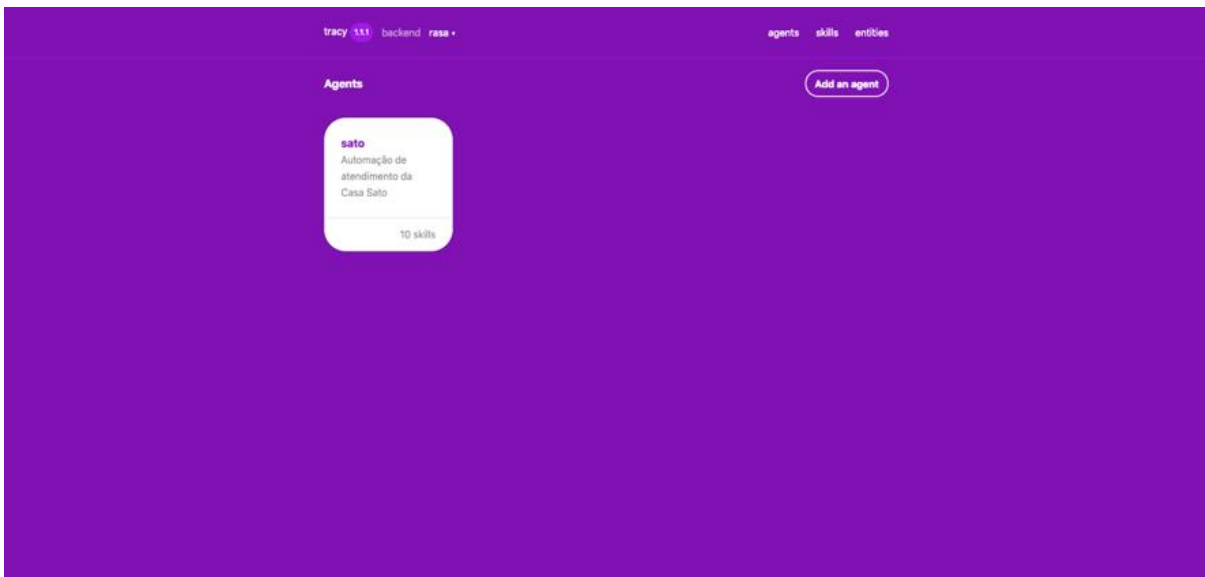
#### 4.3.4 Modelagem de intenções do agente

Para modelagem do agente, foi utilizada a ferramenta Tracy, uma ferramenta visual que auxilia no processo de organização das intenções e estrutura do *chatbot* de forma bem simples e intuitiva, As features estão organizadas da seguinte forma:

- Agents: representa o *bot*, o assistente, etc;
- Skills: Habilidades que um ou mais agentes podem possuir;
- Intents: as intenções de cada usuário;
- Entities: são elementos de informação em intenções;

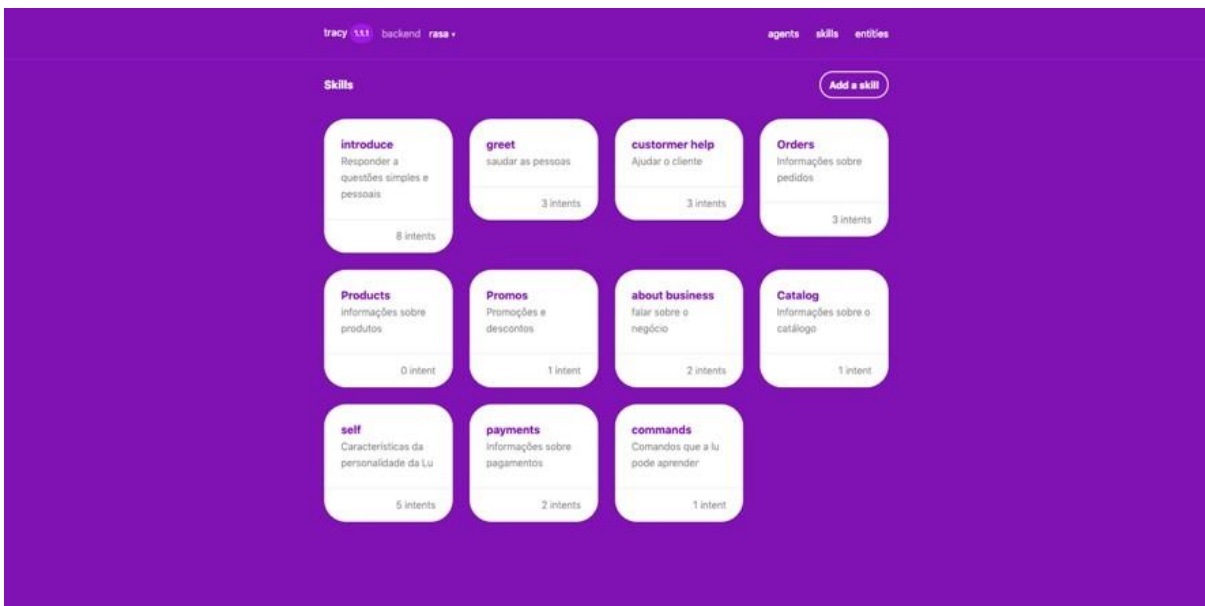
- Slots: definem o que pode ser extraído dos exemplos de intenções.

A figura13, apresenta a definição do agente dentro da ferramenta.



**Figura 13 – Definição do Agente – Tracy**

A figura14, apresenta as Skills, as habilidades que o *bot* possui, dentre elas estão habilidades sobre os traços de personalidade do robô e de relacionamento, como as saudações e despedidas e habilidades voltadas aos problemas apontados no levantamento de dados, como questões sobre a saída de mercadorias, avisos e boletos e novo revendedor.



**Figura 14 – Definição das habilidades (Skills) do bot – Tracy**

Após a definição do agente e das habilidades, é possível fazer o cadastro das intenções respectivas e adicionar exemplos a cada uma, ao final os dados podem ser exportados no formato de treinamento para o Rasa NLU.

### 4.3.5 Estrutura do Chatbot

Dando prosseguimento ao processo de desenvolvimento, após a classificação dos eventos, a construção do diagrama de pareto e as definições do agente, esta seção aborda as definições de construção do Chatbot utilizando o Toolkit Rasa.

Segundo (FRÖHLICH; SOARES,2018), um chatbot depende de duas estruturas: um núcleo de processamento e uma matriz de decisões ou base de conhecimento. A figura15, apresenta a estrutura proposta do chatbot.

**Figura 15 – Estrutura do Chatbot Proposta**



O usuário, ao enviar uma mensagem, independente do canal ao qual possa ser inserida (o projeto está construído visando apenas o Facebook Messenger como canal de comunicação), deve ser encaminhada ao gerenciador de diálogos (Rasa Core), o gerenciador, encaminhará ao motor NLU para classificação da intenção, que devolve ao gerenciador o rank das intenções. A partir das políticas e histórias definidas, o gerenciador irá prever qual ação executar, podendo invocar uma ação customizada no servidor de ações (Action server), buscando informações por api ou ações pré-definidas (utterances). Devolvendo ao gerenciador de diálogos que emite uma resposta, ao usuário respectivo pelo mesmo canal de solicitação.

### 4.3.6 Modelagem de estórias

O Rasa Core, utiliza estórias de usuário, com as quais é possível descrever o fluxo de conversação inicial e quais ações devem ser tomadas de acordo com cada intenção predita. Como exemplo, tomemos a intenção Greetings, a intenção está relacionada a expressões de saudação,

como "oi", "olá", "e ai?", podemos definir uma estória na qual sempre que ocorrer uma intenção Greetings, ele irá retornar uma ação utter\_greet.

Para o projeto, foram escritas estórias de saudação, despedida, sobre a regra de negócio a ser resolvida e características pessoais, para tornar mais humana e natural a interação com o chatbot.

## **4.4 Ambientes**

Como ambiente de testes foi um macbook pro, modelo mid 2012, processador i5 2,5ghz, 16gb de ram DDR3 1333MHz e 240gb de SSD. Para produção, foi utilizada uma VPS (Virtual Private Server), com Ubuntu Linux versão 18.10, processador 1 vCore, 2048MB de RAM e 40gb de SSD.



## 5 RESULTADOS

### 5.1 Testes com usuários

Concluída as etapas de avaliação dos modelos e construção do chatbot, o processo de pesquisa seguiu para o próximo passo: testes com os usuários. Foram selecionados de forma aleatória os clientes da empresa, os testes foram realizados ao longo de 5 dias, com aproximadamente 3 clientes por turno (manhã e tarde), totalizando 30 avaliações. O intuito dos testes foi avaliar aspectos técnicos, para realização de correções e de experiência do usuário, para compreender suas expectativas e como se sentem em relação a tecnologia.

A intenção de avaliar estes pontos é saber se o chatbot irá responder corretamente as indagações dos clientes, se o modelo está prevendo corretamente as intenções e se necessário, realizar correções e ajustes nos modelos a partir das inferências e possíveis erros ocorridos nos testes, possibilitando uma versão de produção mais concisa e validada. De maneira geral, os seguintes itens foram avaliados:

- Facilidade e experiência de atendimento
- Consegui resolver meu problema
- Respostas emitidas pelo *bot*:
  - O robô compreendeu minha pergunta corretamente
  - As respostas do robô era relacionada à pergunta realizada
  - A linguagem utilizada foi adequada e suficientemente clara

O item **facilidade e experiência de atendimento**, pode ser avaliado como muito satisfatório, satisfatório, pouco satisfatório e ruim. A intenção desta avaliação, é conhecer se os usuários conseguiriam interagir sem grandes dificuldades com a plataforma.

O item **Consegui resolver meu problema**, com uma opção com apenas dois valores possíveis: sim ou não. A intenção desta avaliação é conhecer se o chatbot está conseguindo auxiliar a resolver as questões propostas pelos clientes.

Sobre as respostas emitidas pelo *bot*, algumas questões foram levantadas como se o **robô compreendeu pergunta corretamente**, se **a resposta do robô era relacionada a pergunta realizada** e se **a linguagem utilizada foi adequada e suficientemente clara**, cada questão poderia ser avaliada com os valores sim ou não para cada pergunta. A intenção destes questionamentos é verificar se as expressões foram corretamente mapeadas, caso não, mapear as novas expressões (em caso de erro) e verificar se o público alvo estava à vontade com a linguagem utilizada (certas ou erradas) e se as respostas eram suficientemente claras (certas ou erradas). A tabela 10 apresenta os dados coletados nesta etapa.

Avaliação	Facilidade	Experiência
Muito Satisfatório	21	5
Satisfatório	9	12
Pouco Satisfatório	0	8
Ruim	0	5
Total	30	30

**Tabela 10 – Avaliação das Usuários sobre o atendimento realizado pelo Bot**

Analisando os dados da tabela 10, para 70% dos usuários, a facilidade de uso do chatbot foi considerada muito satisfatória, acreditamos que esta avaliação está intrinsecamente atrelada ao fato da plataforma utilizada ser o facebook messenger, por já ser um produto de uso comum e extremamente simples de ser utilizado.

A experiência de interação com o *bot* foi considerada **muito satisfatória** para 16,6% dos usuários, **satisfatória** para 40%, **pouco satisfatória** para 26,6% e **ruim** para 16,6%. Foi qualificaram como satisfatória a facilidade de uso, contudo, a experiência foi considerada positiva apenas para 56,6% dos usuários, para os demais 43,4%, não foi uma experiência positiva. Para melhor compreender as avaliações de experiência, um segundo questionário foi aplicado, com perguntas relacionadas à experiência de uso, o intuito deste questionário é conseguir compreender a razão das avaliações de experiência realizadas, a tabela 11 apresenta os dados de avaliação.

Perguntas	Sim	Não
Robô compreendeu minha pergunta ?	16	14
resposta do robô era relacionada a minha pergunta?	16	14
A linguagem utilizada era adequada?	18	12
Problema foi resolvido?	12	18
Você usaria o bot novamente?	25	5

**Tabela 11 – Experiência com as respostas do Bot**

Analisando os dados da tabela 11, quando questionados se o *bot* conseguiu compreender a pergunta, para 53,3% dos usuários ele conseguiu compreender a pergunta corretamente, o que impacta diretamente na resposta emitida, culminando em respostas igualmente positiva apenas para 53,3% dos usuários. Para 60% dos usuários a linguagem utilizada foi adequada, estava coerente ao contexto e conseguiram compreender bem a mensagem. 40% dos usuários conseguiram resolver seus problemas diretamente com o *bot*, conseqüentemente os demais 60% não conseguiriam resolver e precisariam ser encaminhados a um atendente humano. Quando questionados se utilizariam o *bot* novamente, 16,6% sinalizam que não voltariam a utilizar a tecnologia e 73,4% responderam que sim, que compreendem o momento de teste e avaliam como um iniciativa positiva da empresa na melhoria do atendimento.

Os testes geraram dois produtos para análise: as expressões utilizadas e as definições de fluxo conversacional, ambos geraram algum tipo de impacto. A tabela 12, apresenta as expressões

inseridas durante os testes, a ação esperada, a ação que foi retornada e o problema que pode ser identificado.

Expressão	Ação Esperada	Ação Retornada	Problema
e ai	greet	Fallback	Expressão Não Mapeada
como pego o, catalogo em pdf	catalogo_digital	catalogo	Não identificou a entidade
consulta spc?	não mapeada	não mapeada	Intenção não mapeada
kedi o catalogo	catalogo	Fallback	Expressão não mapeada
Como fasso meu,cadastro?	cadastro	como lucrar	expressão não mapeada e Erro de classificação (Intent Ranking)
Como faço pra.envia os pedido	enviar_pedidos	enviar_pedidos	Resposta mal estruturada
tem como mandar o,catálogo por PDF por favor	catalogo_digital	catalogo	não identificou entidade
eu gostaria de,pegar o catálogo	catalogo	como lucrar	Erro de classificação (Intent Ranking)
po	Fallback	Como Lucrar	Erro de classificação (Intent Ranking) e Predição

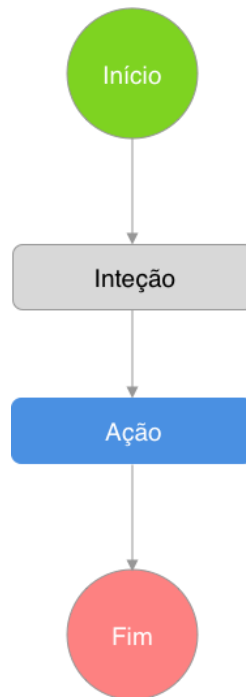
**Tabela 12 – Expressões X Erros**

É possível verificar que ocorreram erros no uso de expressões coloquiais ou erros de grafia, como "e ai", "kedi", "fasso". A primeira versão utilizada nos testes não contou com o mapeamento destas expressões, o segundo ponto observado e a necessidade de intenções complementares para atingir as intenções principais, por exemplo, as expressões "consulta spc?" e "consulta serasa?", foram utilizadas para tirar dúvidas que permeavam ao cadastro, antes do usuário chegar ao seu objetivo, que era o cadastro.

Ainda sobre os dados da tabela12, considerando os erros de classificação, ranqueamento das intenções e identificação de entidades, os erros de grafia e o uso mais coloquial tiveram um impacto neste primeiro teste, devido a não-presença deste tipo de expressão na primeira versão do corpus, ou seja, o algoritmo não consegue avaliar de maneira efetiva. As intenções de catálogo, ocorreram erros de não mapeamento, de classificação, o que podemos considerar uma consequência desse não mapeamento e erros de identificação de subintenções, catálogo, possui catálogos digitais e físicos, tendo subintenções específicas, os erros ocorreram devido ao não reconhecimento das entidades pelo extrator NER presente. O desenho das estórias também foi reavaliado e reestrutura de forma que representasse de maneira mais próxima a interação real que foi realizada durante os testes, a figura16apresenta o primeiro modelo criado.

Existem dois elementos presentes na interação: círculos e retângulos, os círculos definem o início e o fim, sendo o círculo verde o de início da interação e o vermelho indicando o fim. Os retângulos, representação as intenção e as ações de resposta, nas cores cinza e azul, respectivamente. A interação é desenvolvida de forma linear para cada intenção e resposta, a interação é iniciada, o chatbot repassa a expressão ao NLU que retorna com o ranking de intenções classificadas para aquela expressão, o Core analisa o ranking e escolhe a melhor ação a ser executada. O grande entrave desta primeira abordagem é a construção rasa das estórias, não explorando a interação que pode haver entre múltiplas intenções na construção de um diálogo, culminando em poder de predição de diálogos e aprendizado ao Rasa Core bem menor, justamente por não estar construindo a relação entre os modelos de forma mais próxima ao real.

Para contornar este problema e dar mais efetividade a ferramenta, a partir da análise das interações dos usuários de testes, o modelo de estórias foi remodelado, tentando se aproximar



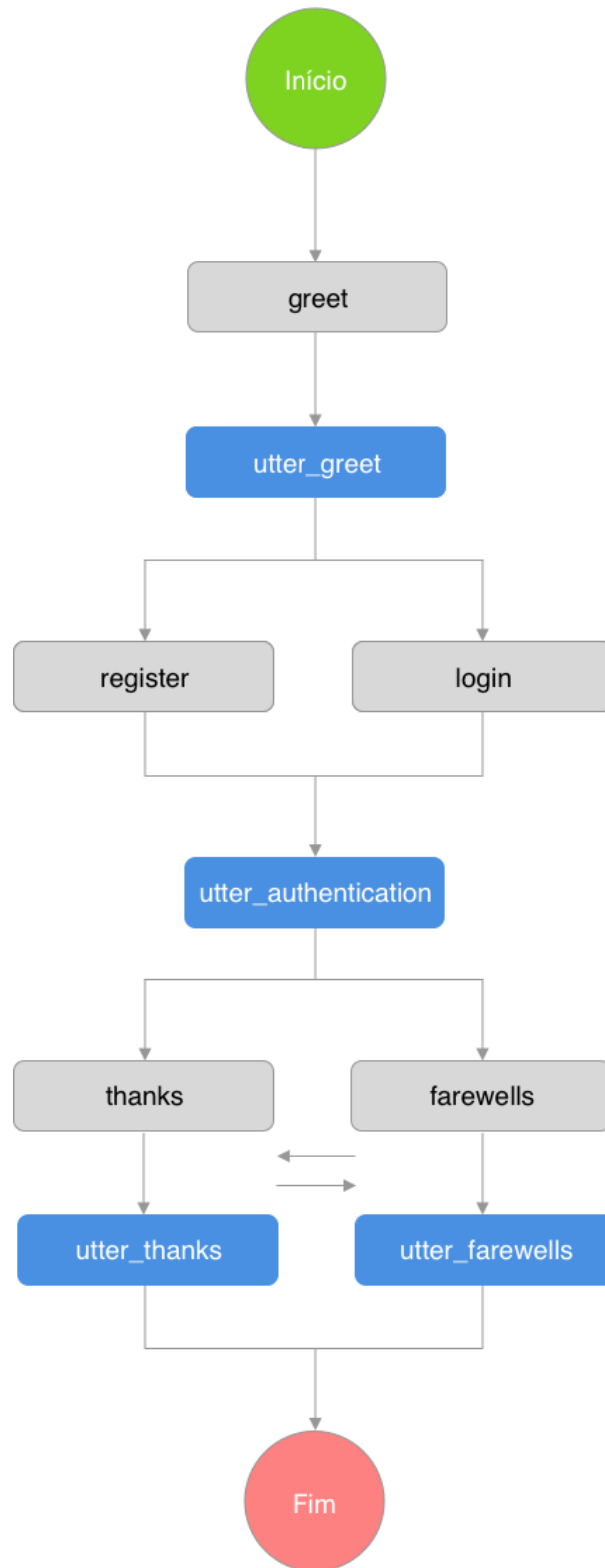
**Figura 16 – Fluxo Conversacional Chatbot**

ao máximo da forma de interação real dos clientes, a figura17, apresenta um dos modelos de estória que foram aproximados.

A figura17, apresenta o modelo de interação para um usuário que deseja realizar cadastro ou login na plataforma, na sua forma mais simples e objetiva. Nele, podemos identificar uma ação de saudação (greet), que retorna uma ação utter\_greet, após a ação, o Rasa Core aguarda a próxima ação do usuário, que neste modelo, entra com uma ação de login ou cadastro que retorna a ação utter\_authentication, após o usuário ter sua resposta, o usuário pode agradecer e se despedir ou vice-versa, finalizando o fluxo desta estória. Foram desenhados fluxos mais complexos para interações maiores, mas é possível compreender a remodelagem das estórias.

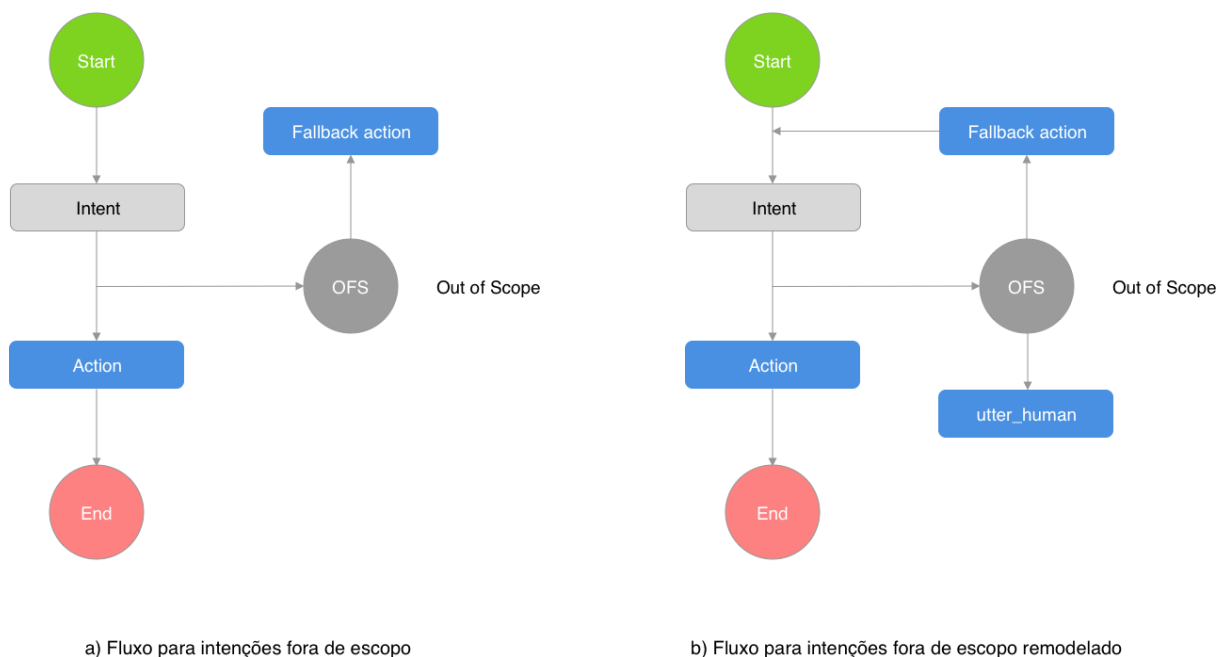
A análise das expressões nos testes resultaram na remodelagem do fluxo e de regras bot nestas ocasiões, por exemplo, um usuário insere expressões que não são conhecidas pelo chatbot repetidamente, o que ele deve fazer? Durante os testes e partir dos erros, foi possível identificar que com a continuidade dos erros, eles apenas exibiam a mensagem definida como resposta para quando não compreendesse a intenção: "desculpe, não entendi". Contudo, em caso de erros repetitivos, a mensagem era repetidamente exibida, sem nenhum feedback ao usuário de como resolver esta situação. A partir dos erros nos testes, um segundo fluxo e regras foram definidas para atender e cobrir esse tipo de erro. A figura18, apresenta o fluxo para as questões fora de escopo (*out of scope*) na versão utilizada nos testes (à esquerda, figura **a**) e na versão remodelada (à direita, figura **b**).

O item **a** da figura18, apresenta a primeira definição para as questões fora de escopo, nele vemos que em caso de expressão fora do escopo definido, o cliente ficaria em um loop



**Figura 17 – Fluxo Conversacional *Chatbot* - remodelado**

voltando para a escuta (action listen), até entrar com uma expressão que fosse conhecida pelo chatbot. Contudo, nos testes foi comprovada a inviabilidade deste modelo, mostrando-se pouco eficaz e não permitindo ao usuário ter outra forma de contactar o atendimento. Para contornar este problema, foi definido um novo fluxo, apresentado no item **b** da figura 18, o modelo possui uma



**Figura 18 – Fluxo conversacional para expressões fora de escopo**

regra que após três tentativas que estejam fora de escopo, ou seja, expressões que o chatbot não consegue classificar ou reconhecer, ele dispara um gatilho que apresenta a ação **utter\_human**, está ação trás ao usuário um questionamento perguntando se ele deseja falar com um humano, dando assim a oportunidade de falar com um humano e contornando o problema.

Finalizada a etapa de testes, as expressões de cada intenção que o bot não conseguiu realizar foram catalogadas e inseridas ao modelo, foram feitas as devidas correções e melhorias e o chatbot foi disponibilizado para interação aos clientes em ambiente de produção.

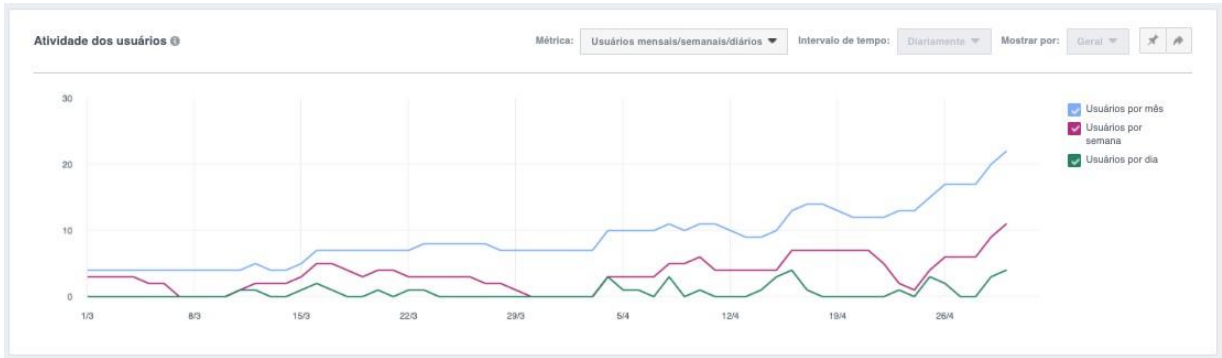
## 5.2 Resultados em Produção

Após a realização dos testes, as expressões e as formas de interação que o chatbot não compreendeu ou não executou corretamente, foram coletadas, corrigidas e adicionadas. Seguindo para o processo de implementação do chatbot em ambiente de produção. O processo de integração é realizado mediante aplicativo dentro da plataforma *Facebook developers*<sup>1</sup>. Todo o processo, deve seguir as diretrizes da plataforma, sendo submetido para aprovação da plataforma e após aprovado, o chatbot é liberado para uso. Doram coletadas informações nos 56 dias seguintes, dentro da plataforma *Facebook developers*.

Neste intervalo, houve interações com 27 usuários, como pode ser acompanhado no gráfico7. Destes atendimentos, foram coletados dados de interação e retenção. O gráfico8, apresenta os 4 tipos de interações do chatbot:

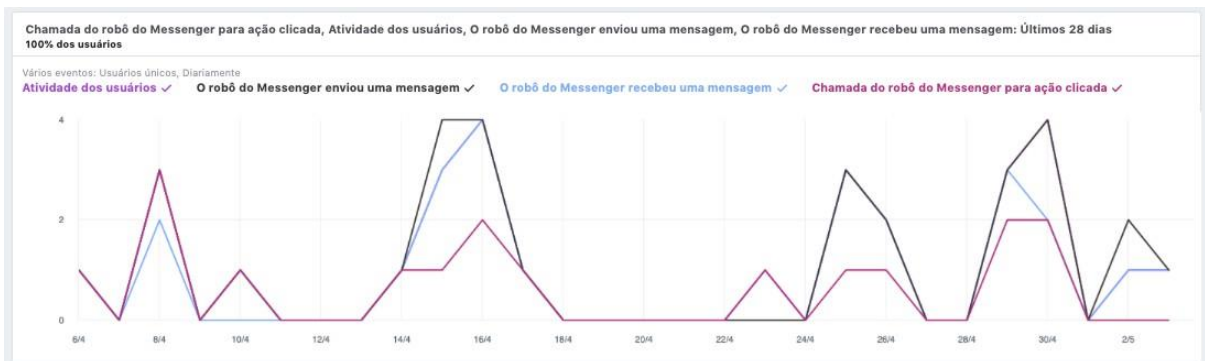
- Atividade dos usuários

<sup>1</sup> <http://developers.facebook.com>



**Gráfico 7 – Gráfico de usuários**

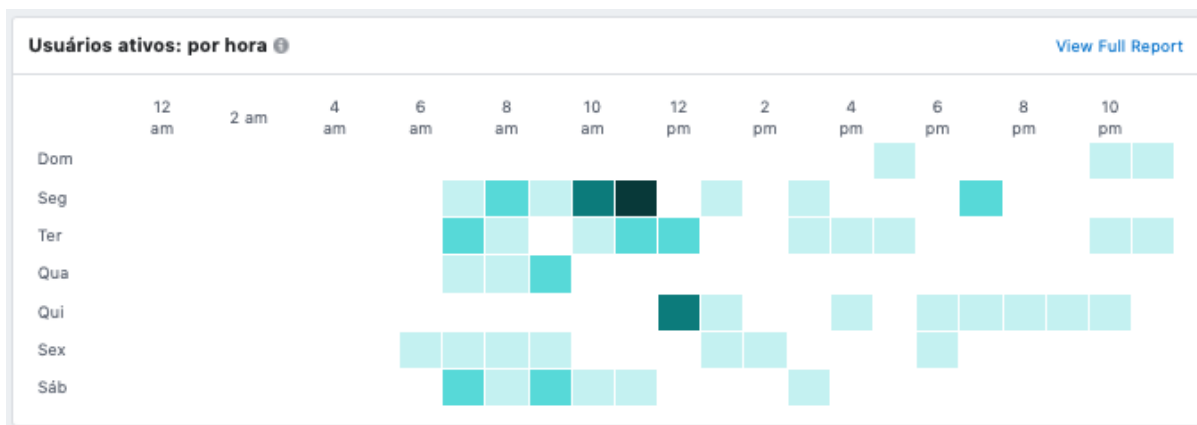
- O robô do Messenger enviou uma mensagem
- O robô do Messenger recebeu uma mensagem
- Chamada do robô do Messenger para ação clicada



**Gráfico 8 – Formas de Interação**

As ações vão do momento 0 (momento inicial) até o momento 4, como pode ser acompanhado no gráfico8, a maioria das interações iniciaram por uma ação clicada e em alguns picos, houve interação de troca de mensagens com o bot. Segundo (PAIVA,2018), 39% das interações desenvolvidas para os chatbots por empresas são guiadas a partir de um roteiro pré-definido. O processo de interação guiada ou botões de ação, auxiliam no processo de decisão do cliente.

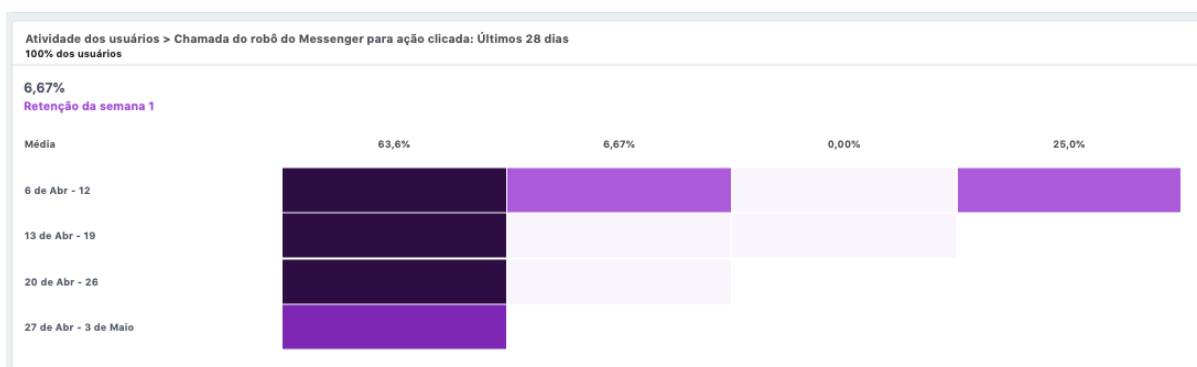
O gráfico9, apresenta os horários de atividade do chatbot. Os tons mais intensos, indicam maior atividade e os mais suaves, menor atividade. Podemos observar que o chatbot esteve em atividade em todos os dias da semana, de 6 horas da manhã até aproximadamente 23 horas. A maior concentração de atividade está nos horários comerciais, e com maior incidência de 10 até às 12 horas. Como pode ser observado, o chatbot pode realizar atendimentos fora de horário comercial da empresa, à partir das 18 horas e aos fins de semana. Este é um aspecto positivo, por permitir o atendimento ao usuário, mesmo em um horário de não-funcionamento da empresa, com um tempo de resposta relativamente curto (1 segundo).



**Gráfico 9 – Horários de Atividade do Chatbot**

O gráfico10, apresenta o percentual de retenção do chatbot, segundo o *Facebook Analytics*<sup>2</sup>. O gráfico apresenta a média de retenção em períodos de 7 dias, dividido em 4 "blocos", no momento inicial de cada semana, é possível verificar a média de retenção é de 63,3%, com mínima de 42% e máxima de 100%. Seguida de 6,67%, 0% e finalizando em 25%.

Podemos concluir que os períodos de retenção mais baixa apresentados no gráfico10, podem estar relacionados aos períodos de menor atividade, como pode ser visualizado no Gráfico 7. Existe uma maior demanda no início de cada período semanal, o que pode ser confirmado em análise do gráfico9, com maior concentração de atividade às segundas e terças-feiras.



**Gráfico 10 – Retenção na Plataforma Messenger**

### 5.2.1 Interações dos clientes em produção

Concluída a etapa de análise de retenção do chatbot, foram analisadas as interações do chatbot com os clientes da empresa, para avaliar e melhor compreender os resultados das mudanças realizadas após os testes. Dos 27 atendimentos realizados, 5 foram separados para análise das expressões, do modelo de interação, da compreensão das expressões e se os usuários conseguiram atingir seu objetivo. A tabela13 apresenta algumas expressões inseridas pelos usuários, e qual a intenção da ação.

<sup>2</sup> <http://developers.facebook.com/analytics>



Expressão	Intenção
Eu não tou com conseguindo mim cadastra	Signup
Posso fazer meu pedido hiroshima online?	Orders
Mas pede uma comta ieu colo do f e não. Da	Out of Scope
Ainda não sou cad	Signup
O que faço pra me cadstra	Signup
Oi quero revender os produtos Hiroshima	Advantages

**Tabela 13 – Expressões dos usuários analisadas**

Analisando a tabela 13, algumas das expressões inseridas não estão na forma culta da língua portuguesa. O primeiro corpus estruturado para testes, utilizava predominantemente a normal culta e após os testes foram inseridas expressões mais coloquiais e informais, buscando maior similaridade com as expressões utilizadas nos testes, ficando mais próximas da linguagem utilizada por grupos de clientes da empresa. Como resultado desta mudança o chatbot conseguiu compreender melhor e entregar aos clientes respostas mais alinhadas as suas intenções, contudo, ainda assim houve a não identificação de algumas expressões, como a expressão "Mas pede uma comta ieu colo do f e não. Da", confirmando a necessidade de evolução contínua.

Complementar a análise das expressões, foram levantados dados do fluxo de conversação, tipos de interação e gatilhos de escopo. A figura 19, apresenta o fluxo conversacional de um usuário para análise mais minuciosa.

Apesar da linearidade apresentada na figura 19, o fluxo convencional esperado era que o usuário primeiro buscava informações sobre o que ele deseja adquirir: o cadastro, catálogo, etc. O que podemos extrair do fluxo é que o usuário inicialmente está solicitando o objetivo final: o cadastro, o catálogo, etc. Depois busca informações adicionais sobre o negócio ou produto, se é necessário realizar algum tipo de investimento inicial (compra de kit), se existe alguma condição (consultas de CPF ou Serasa) e como ele pode lucrar neste modelo de negócio. Essa postura especialmente se mostra interessante por abrir portas para estudos comportamentais e até análise mais aprofundada do perfil do cliente e das interações, deixando-as melhores desenhadas a esta forma de interação (conversa guiada). A figura 20, apresenta o modelo de interação dos clientes com o chatbot em produção.

A figura 20, apresenta duas capturas de tela de interações distintas com o chatbot em produção na plataforma de mensagens do Facebook. À esquerda, temos a interação do cliente que busca informações de como lucrar e à direita, a interação do cliente que está buscando informações de como revender ao produto parceiro da empresa. A predominância do uso da conversa guiada: script de atendimento, a partir de um fluxo modelado e contando por botões de ação em intenções chave, dando direcionamento para atingir objetivos ações como cadastro e catálogo.

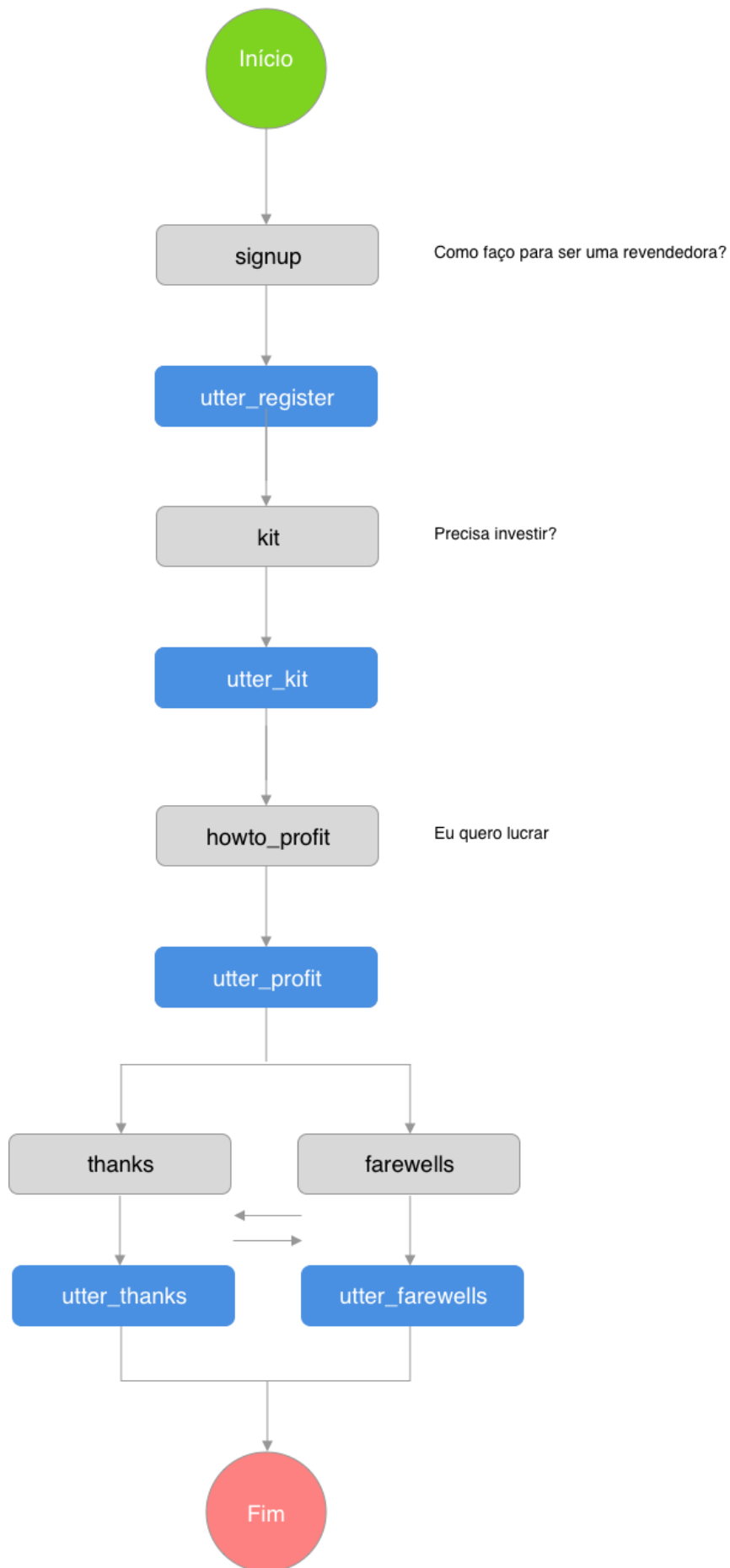
Como apresentado no gráfico 8, houve uma preferência pelo uso da conversa guiada, como apresentado na figura 20. A mudança de abordagem adotada a partir dos erros de testes,

mostram que o uso da conversa guiada e botões, estão sendo bastante utilizados e podemos concluir que está sendo uma ferramenta utilizada pelos clientes. O último ponto analisado, é se a nova abordagem de regras para situações fora de escopo funcionaram corretamente, a figura 21 apresenta uma, dentre às 5 ocorrências na qual o bot trás a opção ao cliente de falar com um humano após três erros consecutivos em não compreender a intenção proposta.

A figura 21, apresenta uma das interações em que o chatbot consecutivamente não conseguiu compreender a expressão inserida e chegou ao limite definido de três erros consecutivos, ativando o fluxo que chama um método que pergunta ao usuário se ele deseja falar com um humano. Esta característica, também foi observadas nos testes e sofreu mudanças para tornar o tratamento aos erros mais amigável ao usuário e fornecer um gatilho que permita que ele saia de um loop contínuo de "desculpe, não entendi" e permitir que os atendimentos que o chatbot não possa realizar, possam ser feitos por humanos, não frustrando o cliente que busca ajuda.

### 5.2.2 Avaliação da equipe de atendimento e vendas

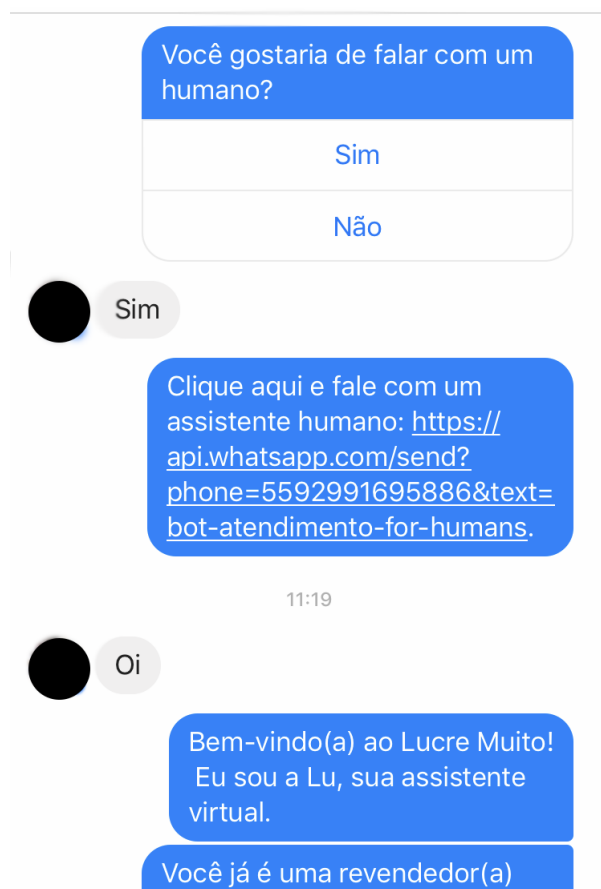
As equipes de atendimento e vendas avaliaram a experiência como positiva, quando questionados se houve alguma resistência ao uso da tecnologia, 40% da equipe afirmou que houve algum tipo de resistência, motivada principalmente pelo receio de serem substituídos pelo *bot* no atendimento ao cliente. Sobre as dificuldades encontradas no processo de elaboração do chatbot, 80% da equipe afirmou que sentiu alguma dificuldade em relação ao processo de construção. Para 60% da equipe, a dificuldade está atribuída ao fato de o *bot* exigir indiretamente a modelagem de um processo de atendimento, o que não havia sido implantado até o momento, para os demais 40% a maior dificuldade ocorreu na implantação de uma FAQ (Frequency Asked Questions), por envolver processo de levantamento, catalogação e elaboração de respostas padronizadas para cada problema catalogado.



**Figura 19 – Análise do fluxo de intenções em produção**



Figura 20 – Interações entre o chatbot e o cliente - Fluxo de conversa guiada



**Figura 21 – Interações entre o chatbot e o cliente - Expressão fora de escopo**

## 6 CONSIDERAÇÕES FINAIS

### 6.1 Discussão

#### 6.1.1 Métodos e Resultados

A pesquisa de (FRÖHLICH; SOARES,2018), utiliza métodos de pesquisa quantitativos e qualitativos, com objetivo conhecer o quão sensível possa estar o usuário considerado não-digital em um contexto de automação de atendimento. O método científico elaborado pelo autor, foi alinhado a diferentes métodos e autores. Em sua pesquisa, os autores buscaram avaliar variáveis de Confiabilidade, como a ausência de erros durante o atendimento; Responsividade, no fornecimento imediato e rápido de atendimento; e Confiança, avaliando se o chatbot possui o conhecimento necessário e consegue compreender as necessidades do cliente, corretamente.

As variáveis por nós definidas foram exploradas de forma diferente, acreditamos que a natureza e o objetivos de pesquisa são diferentes e complementares, considerando os aspectos técnicos avaliados nesta pesquisa e seus direcionamentos para difusão de uso das ferramentas utilizadas que culminam no objeto de pesquisa de trabalhos quantitativos e qualitativos e trabalhos como o de FRÖHLICH; SOARES, nortes para pesquisas técnicas e melhorias ou desenvolvimento de novas ferramentas. Consideramos avaliação por grupos etários e avaliação dos não-usuários ser uma importante contribuição para novas versões do presente trabalho.

As pesquisas de mercado feitas por PAIVA em 2017 e 2018, objetivaram conhecer, definir e quantificar o mercado de *chatbots* no país. A complementariedade dos trabalhos supracitados, devem nortear os direcionamentos futuros dos métodos de pesquisa utilizados aqui, levando em consideração novos aspectos e somando-os aos de mercado citados nos capítulos anteriores, algumas variáveis como o uso de serviços NLU, presentes na pesquisa de PAIVA, com alguns dos principais aspectos validados por FRÖHLICH; SOARES, podem estar relacionadas a disponibilidade de serviço e a confiabilidade, para (BRAUN et al.,2018), a ascensão dos serviços NLU é um dos fatores que contribuíram na popularização de tecnologias conversacionais, como os chatbots e assistentes virtuais.

#### 6.1.2 Tecnologia, Conhecimento e Aplicabilidade

Dentre as tecnologias utilizadas no desenvolvimento de *chatbots*, (SOUZA,2018c), em sua pesquisa, desenvolveu um agente de atendimento voltado para o *e-business*. Em seu projeto, utilizou o Botman<sup>1</sup> como gerenciador de diálogos e o DialogFlow<sup>2</sup>, um serviço de NLU disponível no mercado, como um motor de compreensão e análise, a finalidade do *bot* desenvolvido era responder perguntas frequentes e auxiliar no processo de venda de produtos.

<sup>1</sup> <http://botman.io>

<sup>2</sup> <http://dialogflow.com>

Em um projeto similar, MARCOS; MATTOS, desenvolveu um *chatbot* para auxiliar no processo de atendimento a clientes de um software corporativo, utilizando *IBM Watson*, como solução principal no processamento de intenções de gerenciamento de diálogos.

Ambos os trabalhos optaram pela utilização de utilizar serviços de processamento de intenções, ambos são serviços bem avaliados e presentes no ecossistema de desenvolvimento. (BOCKLISCH et al.,2017), em seu trabalho, apresenta a sua avaliação dos principais serviços NLU do mercado e suas perspectivas sobre o Rasa NLU como alternativa. Acreditamos que o presente trabalho possa somar a este processo, apresentando um estudo de caso que possa ser usado em comparativo aos demais trabalhos.

Podemos concluir a partir dos trabalhos acima citados que, é possível construir um chatbot aplicado aos mais diversos nichos de mercado e de pesquisa, utilizando diferentes tecnologias. Reconhecemos a importância de trabalhos comparando o desempenho de diferentes tecnologias, em igual ambiente de execução, buscando solucionar o mesmo problema de pesquisa. O objetivo desta seção é apresentar a pluralidade no uso deste tipo de tecnologia, sendo aplicada aos mais diversos campos de conhecimento, seja somando esforços no atendimento de clientes, seja no suporte pedagógico de alunos.

## 6.2 Conclusão

Concluimos o presente trabalho com a efetiva implantação do *bot* no canal de atendimento da empresa, dentre os resultados obtidos a retenção de atendimento pode ser destacada. Foi possível reter nos testes até 40% dos atendimentos e em produção, após as devidas correções e ajustes, foi possível obter im retenção média de 63,3%, chegando até máxima de 100% de retenção. Enquadrando-se nas predições de retenção de média para o mercado de (PAIVA,2017) e (FRÖHLICH; SOARES,2018), com uma média de retenção obtida próxima dos resultados de (RODRIGUES,2018) e (MARCOS; MATTOS,2018), em seus estágios iniciais de trabalho, com o diferencial que o presente trabalho de pesquisa utiliza um motor de NLU e Dataset de intenções próprios.

Também foi possível inferir a partir dos resultados de produção, os principais horários de atendimento realizados, identificando atendimentos fora do horário comercial de atendimento da empresa, confirmando aspectos do modelo de atendimento por agentes de (ASPECT,2016), dando disponibilidade de atendimento 7 dias por semana, 24 horas por dia. Por fim, consideramos os resultados positivos, obtendo um agente que pode auxiliar no processo de atendimento, com custo de execução de \$10,00 (dez dólares) ao mês, o que pode ser considerado um custo relativamente baixo. Assim como (RODRIGUES,2018), reconhecemos as limitações dos *chatbots* e corroboramos a importância do agente de atendimento humano.

Assim como (FRÖHLICH; SOARES,2018), acreditamos que o mercado está aberto para novas profissões e na possibilidade de um profissional graduado em Letras atuar no segmento de

tecnologia, como Linguista Computacional. Associado a isso, pontuamos que um processo de melhoria contínua é fundamental para qualidade do *bot*, principalmente relacionado a humanização do robô, alguns obstáculos como a ausência de um processo de atendimento na empresa foram paulatinamente superados, mas a não-definição prévia de um processo foi um fator de impacto negativo para a modelagem do agente.

## **6.3 Trabalhos Futuros**

Para trabalhos futuros, sugerimos a implantação de módulos para checagem de sintaxe e de análise de sentimentos, o desenvolvimento de um corpus de linguagem a partir dos atendimentos realizados. Além de ferramentas de auxílio ao treinamento, manutenção do chatbot, combinação de componentes e teste dos modelos de forma mais rápida e intuitiva.

Realizar um trabalho de estudo do perfil comportamental e linguístico do nicho de clientes e de futuros clientes da empresa, realizando novas adequações e aperfeiçoamentos para que o chatbot possa melhor atender aos clientes da empresa.



## REFERÊNCIAS

- ANDRADE, R. M. Mobile bot: Um chatterbot educacional para dispositivos móveis. **Revista Brasileira de Computação Aplicada**, v. 4, n. 2, p. 83–91, 2013. ISSN 2176-6649.
- ASPECT. 2016 Aspect Consumer Experience Index. **Aspect**, p. 1–17, 2016. Disponível em: <[https://www.aspect.com/globalassets/2016-aspect-consumer-experience-index-survey{\\\_}index-results-final.>](https://www.aspect.com/globalassets/2016-aspect-consumer-experience-index-survey{\_}index-results-final.>)
- ASPECT. **Aspect Agent Experience Index Survey**. [S.l.], 2018. 10 p.
- BASILI, R.; PAZIENZA, M. T.; VELARDI, P. An empirical symbolic approach to natural language processing. **Artificial Intelligence**, v. 85, n. 1, p. 59 – 99, 1996. ISSN 0004-3702. Disponível em:<<http://www.sciencedirect.com/science/article/pii/0004370295001166.>>
- BOCKLISCH, T.; FAULKNER, J.; PAWLOWSKI, N.; NICHOL, A. Rasa: Open Source Language Understanding and Dialogue Management. n. Nips, p. 1–9, 2017. Disponível em: <<http://arxiv.org/abs/1712.05181.>>
- BRAUN, D.; HERNANDEZ-MENDEZ, A.; MATTHES, F.; LANGEN, M. Evaluating Natural Language Understanding Services for Conversational Question Answering Systems. n. August, p. 174–185, 2018.
- CAMBRIA, E.; WHITE, B. **Jumping NLP curves: A review of natural language processing research**. 2014. 48–57 p.
- CHEN, H.; LIU, X.; YIN, D.; TANG, J. A survey on dialogue systems: Recent advances and new frontiers. **SIGKDD Explor. Newsl.**, ACM, New York, NY, USA, v. 19, n. 2, p. 25–35, nov. 2017. ISSN 1931-0145. Disponível em:<<http://doi-acm-org.ez3.periodicos.capes.gov.br/10.1145/3166054.3166058.>>
- CHRISTIANSEN, M. H.; CHATER, N. Connectionist natural language processing: The state of the art. **Cognitive Science**, v. 23, n. 4, p. 417–437, 1999. Disponível em: <[https://onlinelibrary.wiley.com/doi/abs/10.1207/s15516709cog2304\\_2.>](https://onlinelibrary.wiley.com/doi/abs/10.1207/s15516709cog2304_2.>)
- DAMIÃO, M. A.; CAÇADOR, R. M. C.; LIMA, S. M. B. Princípios E Aspectos Sobre Agentes Inteligentes. **Revista Eletrônica da Faculdade Metodista Granbery**, n. 17, p. 1–29, 2014. Disponível em:<<http://re.granbery.edu.br.>>
- FADHIL, A. Domain Specific Design Patterns: Designing For Conversational User Interfaces. 2018. Disponível em:<<http://arxiv.org/abs/1802.09055.>>
- FERREIRA, L. P.; UCHÔA, J. Q. Desenvolvimento de um chatbot para auxiliar o ensino de Espanhol como Língua Estrangeira. **Revista Bazar: Software e Conhecimento Livres**, v. 1, p. 21 – 32, 2006. Disponível em:<<http://repositorio.ufla.br/jspui/handle/1/9629.>>
- FRÖHLICH, L. F. G.; SOARES, V. D. Robotização nos relacionamentos: um estudo sobre o uso de chatbots Robotization. **Revista Fólio**, p. 11, 2018. Disponível em: <<http://www.unisalesiano.edu.br/simposio2013/publicado/artigo0004.pdf.>>
- GIACHOS, I.; PAPAITSOS, E. C.; CHOROZOGLOU, G. Exploring natural language understanding in robotic interfaces. **International Journal of Advances in Intelligent Informatics**, v. 3, n. 1, p. 10, 2017. ISSN 2442-6571.

GIL, A. C. **Como Elaborar Projetos de Pesquisa**. 4. ed. São Paulo: [s.n.], 2002. v. 4. ISSN 85-224-3169-8. ISBN 8522431698.

Grand View Research. **Calcium Carbonate Market Size, Share, Analysis | Industry Report, 2025**. [S.l.], 2016. 110 p. Disponível em: <<https://www.grandviewresearch.com/industry-analysis/chatbot-market>><<https://www.grandviewresearch.com/industry-analysis/calcium-carbonate-market>>.

GURWANI, M. S. **4 Biggest Challenges in Chatbot Development and How to Avoid them**. 2018. Disponível em: <<https://insights.daffodilsw.com/blog/4-biggest-challenges-in-chatbot-development-and-how-to-avoid-them>>.

JÚNIOR, S. R. J. d. S.; BARBOSA, Y. d. A. M. UM CHATBOT PARA RESPONDER FAQs 1. 2017.

KERLY, A.; HALL, P.; BULL, S. Bringing chatbots into education: Towards natural language negotiation of open learner models. **Knowledge-Based Systems**, v. 20, n. 2, p. 177 – 185, 2007. ISSN 0950-7051. AI 2006. Disponível em: <<http://www.sciencedirect.com/science/article/pii/S0950705106001912>>.

KHURANA, D.; KOLI, A.; KHATTER, K.; SINGH, S. Natural Language Processing: State of The Art, Current Trends and Challenges. n. Figure 1, 2017. Disponível em: <<http://arxiv.org/abs/1708.05148>>.

LACSON, R.; KHORASANI, R. Natural Language Processing: The Basics (Part 1). **Journal of the American College of Radiology**, Elsevier Inc., v. 8, n. 6, p. 436–437, 2011. ISSN 15461440. Disponível em: <<http://dx.doi.org/10.1016/j.jacr.2011.04.020>>.

LIDDY, E. D. Natural language processing. **School of Information Studies: Faculty Scholarship School**, p. 67–78, 2014.

LIMA, L. A. Estudo de implementação de um robô de conversação em curso de língua estrangeira em ambiente virtual: um caso de estabilização do Sistema Adaptativo Complexo. p. 131, 2014. Disponível em: <<http://topskills.com.br/wp-content/uploads/2014/11/@TESE-outubro-2014-LUCIANO.pdf>>.

LINDEN, G.; SMITH, B.; YORK, J. Amazon. com Recommendations Item-to-Item Collaborative Filtering. **IEEE Internet Computing**, n. February, 2003.

MANSON, M. **3 types of business chatbots you can build**. 2017. Disponível em: <<https://www.ibm.com/blogs/watson/2017/12/3-types-of-business-chatbots-you-can-build/>>.

MARCOS, A.; MATTOS, R. IBM Watson como Ambiente para Desenvolvimento e Execução de um Chatbot – Um Estudo de Caso Aplicado ao Processo de Atendimento ao Usuário. n. December, 2018.

MORENO, F.; MANFIO, E.; BARBOSA, C. R.; BRANCHER, J. D. Tical: Chatbot sobre o Atlas Linguístico do Brasil no WhatsApp. **Anais do XXVI Simpósio Brasileiro de Informática na Educação (SBIE 2015)**, v. 1, n. Sbie, p. 279, 2015.

NEUMANN, B. Aplicação do ISA e Diagrama de Pareto como ferramentas de gestão do loteamento Lagoa Carapebus Application of isa and pareto diagram as management of the plots Lagoa Carapebus Serra / ES Abstract. n. 2012, 2012.

NORVIG, P.; RUSSEL, S. **Artificial Intelligence A Modern Approach**. [S.l.: s.n.], 2009. v. 3. ISBN 9780136042594.

OLIVEIRA, S. E. D.; ALLORA, V.; SAKAMOTO, F. T. C. Utilização conjunta do método UP (Unidade de Produção -UEP) com o Diagrama de Pareto para identificar as oportunidades de melhoria dos processos de fabricação : um estudo na agroindústria de abate de frango . **Custos e Agronegócio on line**, v. 2, n. 2, p. 37–48, 2006.

PAIVA, F. **Mapa do Ecossistema Brasileiro de Bots 2017**. [S.l.], 2017.

PAIVA, F. **Mapa do Ecossistema Brasileiro de Bots 2018**. [S.l.], 2018. v. 84, 487–492 p. Disponível em:<<http://ir.obihiro.ac.jp/dspace/handle/10322/3933>>.

PEREIRA, F. R.; RIGO, S. J. Utilização de processamento de linguagem natural e ontologias na análise qualitativa de frases curtas. **Novas Tecnologias na Educação**, v. 11, p. 1–10, 2013.

PEREIRA, M. J.; COHEUR, L.; FIALHO, P.; RIBEIRO, R. Chatbots' Greetings to Human-Computer Communication. n. 1994, p. 1–22, 2016. Disponível em:<<http://arxiv.org/abs/1609.06479>>.

RAHMAN, A.; MAMUN, A. A.; ISLAM, A. Programming challenges of chatbot: Current and future prospective. In: . [S.l.: s.n.], 2017. p. 75–78.

RASA. **Rasa NLU: Language Understanding for chatbots and AI assistants**. 2018. Disponível em:<<https://rasa.com/docs/nlu/>>.

RASA. **The Rasa Core dialogue engine**. 2018. Disponível em:<<https://rasa.com/docs/core/>>.

RODRIGUES, T. Enhancing chat bots using machine learning. 2018.

SAMEERA, A. A.-K. S. o. C. S.; IRAQI, E. E. of E. C. D. U. D.; WOODS, J. S. o. C. S.; UK, E. E. of E. C. Review on Chatbot Design Techniques in Speech Conversation Systems. **International Journal of Advanced Computer Science and Applications**, v. 6, n. 7, p. 37–46, 2015. ISSN 23941588.

SHAWAR, B. A.; ATWELL, E. Using dialogue corpora to train a chatbot. In: . [S.l.: s.n.], 2003.

SHUM, H.; HE, X.; LI, D. From eliza to xiaoice: Challenges and opportunities with social chatbots. **CoRR**, abs/1801.01957, 2018. Disponível em:<<http://arxiv.org/abs/1801.01957>>.

SOUZA, R. F. **DESENVOLVIMENTO DE CHATBOT APLICADO AO ATENDIMENTO DE CLIENTES EM E-BUSINESS**. Tese (Bachelor's Thesis) — UNIVERSIDADE FEDERAL DE SANTA CATARINA, 2018. Disponível em:<<https://www.uam.es/gruposinv/meva/publicacionesjesus/capitulos{ }espanyol{ }jesus/2005{ }motivacionparaelaprendizajePerspectivaalumnos.pdf{ }0Ahttps://www.researchgate.net/profile/Juan{ }Aparicio7/publication/253571379{ }Los{ }estudios{ }sob>>.

SOUZA, R. F. **Desenvolvimento de Chatbot Aplicado ao Atendimento de Clientes em E-business**. Tese (Doutorado) — Universidade Federal de Santa Catarina, 2018. Disponível em:<<http://e-journal.uajy.ac.id/14649/1/JURNAL.pdf>>.

SOUZA, R. F. **Desenvolvimento de Chatbot Aplicado ao Atendimento de Clientes em E-business**. 2018. Disponível em:<<http://e-journal.uajy.ac.id/14649/1/JURNAL.pdf>>.

STENT, A.; PRASAD, R.; WALKER, M. Trainable sentence planning for complex information presentation in spoken dialog systems. p. 79–es, 2007.

TURING, A. M. Computing machinery and intelligence. In: \_\_\_\_\_. [S.l.: s.n.], 1950. v. 59.

Villmann, T.; Der, R.; Martinetz, T. A new quantitative measure of topology preservation in kohonen's feature maps. In: **Proceedings of 1994 IEEE International Conference on Neural Networks (ICNN'94)**. [S.l.: s.n.], 1994. v. 2, p. 645–648 vol.2.

WEIZENBAUM, J. Eliza—a computer program for the study of natural language communication between man and machine. **Commun. ACM**, ACM, New York, NY, USA, v. 9, n. 1, p. 36–45, jan. 1966. ISSN 0001-0782. Disponível em: <http://doi.acm.org/10.1145/365153.365168>.