

UNIVERSIDADE FEDERAL DE MINAS GERAIS

Instituto De Ciências Exatas

Departamento De Estatística

Programa de Pós-graduação em Estatística

DANIEL MARTINEZ

**ESTUDO COMPARATIVO ENTRE A CAPACIDADE DE GENERALIZAÇÃO DE
MODELOS CRIADOS A PARTIR DAS PRINCIPAIS PLATAFORMAS DE
DESENVOLVIMENTO DE ASSISTENTES VIRTUAIS**

Belo Horizonte

2021

Daniel Martinez

**ESTUDO COMPARATIVO ENTRE A CAPACIDADE DE GENERALIZAÇÃO DE
MODELOS CRIADOS A PARTIR DAS PRINCIPAIS PLATAFORMAS DE
DESENVOLVIMENTO DE ASSISTENTES VIRTUAIS**

Monografia apresentada ao Programa de Pós-Graduação em Estatística do Instituto de Ciências Exatas da Universidade Federal de Minas Gerais como requisito parcial para obtenção do grau de Especialista em Estatística com Ênfase em Indústria e Mercado.

Orientadora: Profa. Dra. Ilka Afonso Reis

Coorientadora: Profa. Izabella Bauer

Belo Horizonte

2021

2021, Daniel Felipe Martinez do Nascimento.
Todos os direitos reservados

:

	Nascimento, Daniel Felipe Martinez do.
N244e	Estudo comparativo entre a capacidade de generalização de modelos criados a partir das principais plataformas de desenvolvimento de assistentes virtuais [manuscrito]. / Daniel Felipe Martinez do Nascimento — 2021. 41 f. il. Orientadora: Profa. Ilka Afonso Reis Coorientadora: Profa. Izabella Bauer Monografia (especialização) - Universidade Federal de Minas Gerais, Instituto de Ciências Exatas, Departamento de Estatística. Referências 40-41. 1. Estatística. 2. Assistentes virtuais. 3. Chatbot. 4. I. Reis, Ilka Afonso II. Bauer, Izabella. III. Universidade Federal de Minas Gerais, Instituto de Ciências Exatas, Departamento de Estatística .IV.Título.
	CDU 519.2 (043)

Ficha catalográfica elaborada pela bibliotecária Belkiz Inez Rezende Costa CRB 6ª Região nº 1510



Universidade Federal de Minas Gerais
Instituto de Ciências Exatas
Departamento de Estatística
Programa de Pós-Graduação / Especialização
 Av. Pres. Antônio Carlos, 6627 - Pampulha
 31270-901 – Belo Horizonte – MG

E-mail: pgest@ufmg.br
 Tel: 3409-5923 – FAX: 3409-5924

**ATA DO 223ª. TRABALHO DE FIM DE CURSO DE ESPECIALIZAÇÃO EM
 ESTATÍSTICA DE DANIEL FELIPE MARTINEZ DO NASCIMENTO.**

Aos vinte e um dias do mês de maio de 2021, às 14:00 horas, com utilização de recursos de videoconferência a distância, reuniram-se os professores abaixo relacionados, formando a Comissão Examinadora homologada pela Comissão do Curso de Especialização em Estatística, para julgar a apresentação do trabalho de fim de curso do aluno **Daniel Felipe Martinez do Nascimento**, intitulado: “*Estudo comparativo entre a capacidade de generalização de modelos criados a partir das principais plataformas de desenvolvimento de assistentes virtuais*”, como requisito para obtenção do Grau de Especialista em Estatística. Abrindo a sessão, a Presidente da Comissão, Professora Ilka Afonso Reis – Orientadora, após dar conhecimento aos presentes do teor das normas regulamentares, passou a palavra ao candidato para apresentação de seu trabalho. Seguiu-se a arguição pelos examinadores com a respectiva defesa do candidato. Após a defesa, os membros da banca examinadora reuniram-se sem a presença do candidato e do público, para julgamento e expedição do resultado final. Foi atribuída a seguinte indicação: o candidato foi considerado Aprovado condicional às modificações sugeridas pela banca examinadora no prazo de 30 dias a partir da data de hoje por unanimidade. O resultado final foi comunicado publicamente ao candidato pela Presidente da Comissão. Nada mais havendo a tratar, a Presidente encerrou a reunião e lavrou a presente Ata, que será assinada por todos os membros participantes da banca examinadora. Belo Horizonte, 21 de maio de 2021.

Prof.^a Ilka Afonso Reis (Orientadora)
 Departamento de Estatística / ICEX / UFMG

Prof.^a Izabella Bauer de Assis Cunha (Coorientadora)
 Pontifícia Universidade Católica de Minas Gerais

Prof.^a Edna Afonso Reis
 Departamento de Estatística / ICEX / UFMG

Felipe Cruz Neiva Campos
 Google Brasil



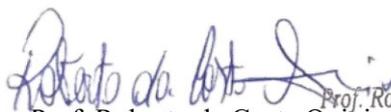
Universidade Federal de Minas Gerais
Instituto de Ciências Exatas
Departamento de Estatística
Programa de Pós-Graduação / Especialização
Av. Pres. Antônio Carlos, 6627 - Pampulha
31270-901 – Belo Horizonte – MG

E-mail: pgest@ufmg.br
Tel: 3409-5923 – FAX: 3409-5924

DECLARAÇÃO DE CUMPRIMENTO DE REQUISITOS PARA CONCLUSÃO DO CURSO DE ESPECIALIZAÇÃO EM ESTATÍSTICA.

Declaro para os devidos fins que **Daniel Felipe Martinez do Nascimento**, número de registro 2017705556, cumpriu todos os requisitos necessários para conclusão do curso de Especialização em Estatística, entregando a versão final do trabalho de conclusão de curso para sua orientadora, a professora Ilka Afonso Reis, que aprovou a versão final. O trabalho foi apresentado no dia 21 de maio de 2021 com o título “*Estudo comparativo entre a capacidade de generalização de modelos criados a partir das principais plataformas de desenvolvimento de assistentes virtuais*”.

Belo Horizonte, 25 de maio de 2021


Prof. Roberto da Costa Quinino
Coordenador do curso de Especialização em Estatística
Departamento de Estatística / UFMG

AGRADECIMENTOS

“Se eu vi mais longe, foi por estar sobre ombros de gigantes.” essa citação de Isaac Newton (1643 – 1727) já nos mostrava a dependência de trabalhos anteriores, da característica evolutiva e incremental da ciência, por isso agradeço aquelas pessoas que dedicam suas vidas a contribuir com a evolução da ciência, muitas vezes de forma quase filantrópica.

Agradeço a minha esposa Renata, que suportou dias afincos ao meu lado, sem que eu pudesse dedicar a atenção que ela merecera sem dúvidas.

Agradeço em especial a Profª. Doutora Ilka Afonso Reis, que sem a qual esse trabalho não seria concluído.

À Profª. Izabella Bauer, que tanto me apoiou, mostrando o caminho e traçando os planos para que o trabalho fosse exequível.

À minha mãe Vilma Martinez e irmã Ágata Martinez, que nos momentos difíceis em que quase desisti me trouxeram refrigério.

E principalmente a Deus, Romanos 11:36 “Portanto Dele, por Ele e para Ele são todas as coisas. A Ele seja a glória perpetuamente!”

“I propose to consider the question, ' Can machines think?’” (Alan Turing, Computing Machinery and Intelligence, 1950).

RESUMO

Chatbots, agentes conversacionais, assistentes virtuais ou mesmo bots são termos utilizados em muitos casos para denominar programas de computador com a capacidade de interação com ser humano em um fluxo conversacional, seja por botões em árvores de diálogos ou, nos mais elaborados, compreendendo linguagem natural. A cada ano a utilização desse tipo de solução é mais comum nas empresas, tanto para públicos externos, como clientes ou fornecedores, quanto para os próprios colaboradores. No entanto, há uma dificuldade que ronda as organizações no momento de decidir sobre qual plataforma utilizar para desenvolver seu chatbot: existe diferença entre os motores cognitivos das principais plataformas? Este trabalho busca responder se, existe ou não diferença entre os motores cognitivos das três plataformas de desenvolvimento de chatbots selecionadas: Alexa Console Developer da Amazon, Dialogflow do Google e Watson Assistant da IBM. O experimento fatorial desenhado neste estudo, comparou a capacidade de generalização dos modelos de Machine Learning para o Português brasileiro das três plataformas, avaliando cinco intenções criadas em cada, dividida em dois conjuntos, genéricas e específicas. Após a criação das intenções, foram avaliadas as classificações de interações comparando ao padrão ouro (*ground truth*), e depois verificada a diferença entre performances de cada plataforma, de forma geral e segregada, ajustado um modelo de regressão logística e calculado a razão de chance (*odds ratio*). A análise dos resultados do experimento mostrou que o Watson Assistant da IBMs tem melhor performance geral, com um percentual de acerto médio de 82%, e com uma odds de acerto 4,43 vezes maior que a da Dialogflow e 3,03 vezes da Alexa Console Developer. Entretanto, em alguns conjuntos segregados, não houve diferença estatística significativa entre as três plataformas, como no Fator Classe da Intenção ao nível Específica. É sabido que o fator desempenho não deve ser o único a ser considerado e fatores como custos, expertises, facilidade na implementação, disponibilidade e integrações, entre outros, também devem ser levados em conta. Porém, como conclusão desse trabalho, foi observada uma diferença significativa entre os motores cognitivos das plataformas avaliadas.

Palavras chaves: Assistentes Virtuais, Chatbots, Dialogflow, Watson Assistant, Alexa Console Developer, Motor Cognitivo, Experimento fatorial, Machine Learning, NLP, Natural Language Processing.

ABSTRACT

Chatbots, conversational agents, virtual assistants, or simply bots are used in many cases to represent computer programs with the ability to interact with human beings in a conversational flow, whether by buttons on decision trees or in other more complex cases with the understanding of natural language. Each year, this solution becomes more popular in companies, whether for external audiences such as customers or employees themselves. However, there is a difficulty around organizations when deciding on which platform to use to develop their chatbot. a question arises: is there any difference between the cognitive engines of the main platforms? This work seeks to answer whether there is any difference between the cognitive engines of the three selected chatbots development platforms, Alexa Console Developer from Amazon, Dialogflow from Google, and Watson Assistant from IBM. The factorial experiment compared the ability to generalize Machine Learning models in Brazilian Portuguese for the three platforms, evaluating 5 intentions created in each, divided into two sets, generic and specific. After creating the intentions, the interaction classifications were evaluated by comparing them to the ground truth, and then the difference between performances of each platform, in general, and separated, was verified, adjusting it to a logistic regression model and calculating the Odds Ratio. The experiment showed interesting results, showing that IBM's Watson Assistant has better overall performance, with an average hit percentage of 82% and an Odds Ratio of 4.43 times higher than Dialogflow and 3.03 Alexa Console Developer. However, there was no statistically significant difference in some segregated settings, as in the Class Factor of Intention at the Specific level, where no difference was significant. It is known that the performance factor should not be the only one to be considered and factors such as costs, expertise, ease of implementation, availability, and integration, among others, must also be taken into account. However, as a conclusion of this work, a significant difference was observed between the cognitive engines of the evaluated platforms.

Keywords: Virtua Assistants, Chatbots, Dialogflow, Watson Assistant, Alexa Console Developer, Cognitive Engine, Factorial Experiment, Machine Learning, NLP, Natural Language Processing.

SUMÁRIO

	AGRADECIMENTOS	3
1	INTRODUÇÃO	11
2	OBJETIVOS	13
	2.1 Objetivo principal	13
	2.2 Objetivo específicos	13
3	PROBLEMATIZAÇÃO	13
4	JUSTIFICATIVA	14
5	REFERENCIAL TEÓRICO	14
	5.1 Inteligência Artificial	14
	5.1.1 Natural Language Processing (NLP)	16
	5.2 Assistentes Virtuais	18
6	METODOLOGIA	22
	6.1 Variável Resposta	22
	6.2 Criação de intenções	22
	6.3 Criando o conjunto de treino e teste	25
	6.4 Classificando em cada plataforma	26
	6.5 Hipóteses de trabalho	29
	6.6 Regressão logística	29
	6.7 Modelo de regressão logística para os fatores plataformas e classe de intenção	31
	6.8 Modelo de regressão logística para os fatores plataforma e tipo de intenção	31
7	RESULTADOS	32
	7.1 Análise descritiva dos dados	32
	7.2 Resultados para cada um dos fatores	35
	7.3 Análise múltipla	37
	7.3.1 Modelo múltiplo – Interação entre classe de intenção e plataforma	37
	7.3.2 Modelo simples – Segregado pelo nível genérico para o fator classe de intenção	37
	7.3.3 Modelo simples – Segregado pelo nível específico para o fator classe intenção	38
	7.3.4 Modelo simples – Comparação entre plataformas para cada nível do fator tipo de intenção	38
8	CONCLUSÕES	41
9	REFERÊNCIAS	43

1 INTRODUÇÃO

A consultoria Gartner apresentou em sua pesquisa anual, denominada *Hype Cycle* de Inteligência Artificial (2019), o número de empresas que implementaram algum tipo de Inteligência Artificial (IA) no mundo, sendo que em 2018 cresceu para 4%, e em 2019 para 14%. Apontou também a IA de conversação como principal promotor de iniciativas para adoção de implementações nas corporações, seguindo a tendência de sucesso dos Assistentes Virtuais como, Alexa, Google Assistant, Siri, Cortana, entre outros.

Apesar de ser um campo de pesquisa com quase 70 anos, a IA ganhou notoriedade nos últimos anos. Em pesquisa pelo Google Trends sobre o termo “Inteligência Artificial” nos últimos 5 anos no Brasil, o pico de busca foi entre os dias 22 e 28 de novembro de 2020.

Percebe-se na citação de Simon da década de 50, a previsão de utilização de IA com capacidade semelhante a mente humana:

Não é meu objetivo surpreendê-lo ou chocá-lo - mas a maneira mais simples de resumir é dizer que agora existem no mundo máquinas que pensam, aprendem e criam. Além disso, sua capacidade de fazer essas coisas aumentará rapidamente até que, em um futuro visível, a gama de problemas que eles possam lidar seja coextensiva à faixa em que a mente humana foi aplicada (HEBERT SIMON, 1957, p. 163).

Uma definição de IA aceita por muitos praticantes dessa arte é a de Marvin Minsky, do MIT: “A IA é a ciência de fazer máquinas fazerem coisas que exigiriam inteligência se fossem feitas por homens.” (DANIEL CREVIER, 1993).

Para Rich and Knight (1991), o conceito de IA aproxima-se a um “estudo de como os computadores podem fazer tarefas que hoje são mais bem desempenhadas pelas pessoas.”, ou ainda “a arte de criar máquinas que executam funções que exigem inteligência quando executadas por pessoas.” (Kurzweil, 1990).

Para Norvig e Russell (2013) a definição de IA pode ser dividida em 4 categorias: “Pensando como Humano”, “Pensando Racionalmente”, “Agindo como humano” e “Agindo racionalmente”. A categoria “Agindo como humano” pode-se referir ao conceito básico dos Assistentes Virtuais, que são máquinas capazes de interagirem com os seres humanos por meio de linguagem natural, com o objetivo de executar uma tarefa, ou mesmo responder a uma solicitação.

Chatbots, Assistentes Virtuais, ou simplesmente *bots*, são programas de computadores que interagem com usuários por uma interface conversacional, seja ela voz ou

texto, processando a entrada em linguagem natural e entregando uma saída ao usuário como uma resposta. Um dos primeiros *bots* a entrarem em produção em larga escala no Brasil foi a BIA (Bradesco Inteligência Artificial), trazendo uma experiência diferente para os clientes do banco. Com uma estratégia diferente, o *bot* da Magazine Luiza, denominado Lu, passou a ser uma *Influencer* Digital, e não somente um *bot*. O Assistente do Banco do Brasil foi responsável por mais de um milhão de atendimentos em 2019 ¹. Já a MRV adotou um nome compatível com seu acrônimo, a Maria Rosa Vaz, seu *bot* desde 2017 ², que em 2020 passou a se chamar Mia (MRV Inteligência Artificial).

Essa popularização na utilização e desenvolvimento de *bots* aumentou no último ano consideravelmente. A pesquisa do Panorama *Mobiletime* (2020) do ecossistema brasileiro de *bots* apresentou um aumento de 17 mil *bots* para 60 mil em 2019, representando um crescimento de 252%, e de 60 mil em 2019 para 101 mil em 2020, segundo empresas respondentes da pesquisa.

A base para o desenvolvimento de um Assistente Virtual Inteligente é a utilização de NLP (*Natural Language Process*). Por meio da NLP, um Assistente Virtual consegue entender o que está sendo dito e interagir com um humano. Uma definição sobre NLP pode ser encontrada no site da IBM como:

O processamento da linguagem natural (NLP) é uma área da inteligência artificial que utiliza a linguística computacional que fornece análise e interpretação semântica do texto, o que permite aos sistemas aprender, analisar e entender a linguagem humana (IBM, 2020).

Segundo o Prof. EISENSTEIN Jacob, NLP é o conjunto de métodos que tornam possível o acesso à linguagem humana pelos computadores. Esse acesso é fundamental para o desenvolvimento de um Assistente Virtual Inteligente.

Apesar de alguns autores defenderem a diferença entre *Chatbots* e Assistentes Virtuais, trataremos nesse trabalho ambos como *Chatbots* ou simplesmente *bots*.

Para entendermos como os *Chatbots* conseguem interagir com as pessoas, se faz necessário a compreensão sobre semântica, que é conceituado como “o ramo da linguística voltado para a investigação do significado das sentenças” (CANÇADO, 2008).

No contexto deste trabalho foi utilizado o termo Análise Semântica como a capacidade da plataforma de construção de *Chatbots* de extrair significado sobre as frases inseridas como exemplo nas intenções, sendo as intenções os objetivos ou propósitos expressos em uma entrada do usuário.

Com efeito, neste trabalho, foi realizada uma pesquisa experimental entre plataformas de desenvolvimento para Assistentes Virtuais, com o objetivo de responder à seguinte questão de pesquisa: Existe diferença entre o motor cognitivo para intenções das plataformas de desenvolvimento de assistentes virtuais em português/Brasil, Dialogflow da Google, Watson Assistant da IBM e Alexa da Amazon?

Cada uma das plataformas de desenvolvimento tem suas características específicas, com *features* e desempenho diferentes. O motor cognitivo para intenções é um dos pontos em comum entre todas as plataformas a serem analisadas neste trabalho.

2 OBJETIVOS

2.1 Objetivo principal

Comparar a capacidade cognitiva semântica das intenções entre as plataformas de desenvolvimento de Assistentes Virtuais, Dialogflow do Google, Watson Assistant IBM e Alexa da Amazon, para português do Brasil.

2.2 Objetivo específicos

- Verificar se existe diferença estatística entre o entendimento semântico das intenções entre as plataformas Dialogflow do Google, Watson Assistant IBM e Alexa da Amazon;
- Comparar as classificações por intenções das três plataformas;

Caso exista diferença entre as plataformas, criar um rank das classificações comparado ao padrão ouro classificado pelo humano.

3 PROBLEMATIZAÇÃO

Consultorias como Capgemini, Deloitte e Gartner estimam em seus relatórios a expansão na utilização de *bots* para os próximos anos, incluindo a iniciativa de novos desenvolvimentos. Nessas situações, as empresas precisarão escolher uma plataforma para esse desenvolvimento, caso optem por construir seus próprios *bots*. Entre as variáveis existentes como expertise na plataforma, preço e facilidade de integrações, existe uma dúvida que permeia essa escolha: será que existem diferenças entre os motores cognitivos das principais plataformas de desenvolvimento de *bots*?

Esse trabalho espera minimizar o peso dessa variável na equação, que poderá auxiliar na escolha da melhor plataforma no contexto de cada empresa.

4 JUSTIFICATIVA

Empresas estão começando a visualizar o valor da utilização de *chatbots*, com oportunidades de redução de custos de atendimentos, ampliação do horário disponível ao público 24h por dia e 7 dias por semana, mantendo um atendimento uniforme e alinhado com o discurso da marca.

A utilização de ferramentas para desenvolvimento de Assistentes Virtuais vem crescendo, com uma perspectiva de expansão ainda maior para os próximos anos, e isso justifica a quantidade de novos ambientes para desenvolvimento que vêm sendo lançados. Grandes players como Amazon, Google, IBM e Oracle já possuem plataformas proprietárias.

Portanto, a escolha de qual plataforma utilizar para a construção do seu *chatbot* se torna ainda mais desafiador, principalmente para empresas que possuem pouca expertise na área.

5 REFERENCIAL TEÓRICO

5.1 Inteligência Artificial

O surgimento da IA ocorreu na década de 30 (Barr & Feigenbaum, 1981), mas foi em 1956 que teve seu nome cunhado oficialmente como inteligência artificial em uma conferência de verão em Dartmouth College, NH, USA (McCorduck, 1979).

Ao longo das décadas pode-se perceber a evolução da IA entre vales e picos, sendo o período de 1952 a 1969 considerado um período de grande entusiasmo (Norvig e Russell, 2013), com a criação em 1959 do Geometry Theorem Prover, um dos primeiros programas de IA.

Na década seguinte, de 1966 a 1974, a IA passava por uma transformação e era utilizada quase que exclusivamente no campo acadêmico, com formulação matemática e critérios mais rigorosos para ser considerada uma aplicação de IA. Foi neste período que os primeiros sistemas especialistas foram criados, sendo um deles, um sistema de detecção de infecção no sangue com mais de 450 regras (Norvig e Russell, 2013).

Já na década de 80 o Japão lança o projeto Quinta Geração, cujo objetivo era criar um computador inteligente capaz de realizar milhares de inferências por segundo. Outro objetivo do projeto era desenvolver programas que fossem capazes de entender a linguagem natural.

Também foi nessa década que as redes neurais retornaram a ganhar destaque (Norvig e Russell, 2013).

A década de 90 é considerada até a atualidade a era da IA moderna, na qual houve grande salto com programas capazes de compreender linguagem natural, guiar carros autônomos, prever comportamentos, recomendar produtos entre outras diversas aplicações.

Em 1997 o DeepBlue, desenvolvido pela IBM, ganha uma partida de xadrez do até então atual campeão mundial desse esporte Garry Kasparov. Em fevereiro de 2011 o Watson da IBM ganha notoriedade ao ganhar o Jeopardy, um programa de televisão americano de perguntas e respostas. Em 2016 foi a vez do Google com o projeto AlphaGO ganhar notoriedade, quando seu programa ganha do Campeão Mundial de GO Lee SEDOL, sendo GO um jogo complexo com uma estimativa mínima de combinações 2×10^{170} (Tromp e Farneback , 2006).

Hoje em 2021 existem diversas aplicações que utilizam Inteligência Artificial (IA). Em um artigo publicado em 22 de Junho de 2020 com o título “10 Wonderful Examples Of Using Artificial Intelligence (AI) For Good”, a revista Forbes elencou 10 exemplos maravilhosos de utilização de IA para o bem e entre eles estão análises de imagens médicas para detecção de câncer, o projeto World Bee Project que colhe e analisa dados de abelhas com o objetivo de criar correlações sobre sua saúde e o meio ambiente para criar insights e assim ajudar a informar e implementar ações globais para melhorar os habitats dos polinizadores; pesquisadores do Instituto de Algoritmos de Aprendizagem de Montreal (MILA) usam GANs (redes adversárias generativas) para simular o dano de tempestades severas e aumento do nível do mar, conservação da vida selvagem, combate à fome e desigualdade no mundo, entre outras aplicações. (Site Forbes, Junho de 2020)

Apesar da ampla utilização da IA, ainda não há consenso entre os principais autores sobre uma definição exata. O conceito de IA se trata de uma discussão filosófica, e existem várias tentativas, como em Norvig e Russell (2013), que divide o conceito em 4 categorias, conforme Quadro 1.

Quadro 1: Definições de IA – Adaptada de Norvig e Russell (2013)

Pensando como um humano	Pensando racionalmente
<p>“O novo e interessante esforço para fazer os computadores pensarem (...) <i>máquinas com mentes</i>, no sentido total e literal.” (Haugeland, 1985) “[Automatização de] atividades que associamos ao pensamento humano, atividades como a tomada de decisões, a resolução de problemas, o aprendizado...” (Bellman, 1978)</p>	<p>“O estudo das faculdades mentais pelo uso de modelos computacionais.” (Charniak e McDermott, 1985) “O estudo das computações que tornam possível perceber, raciocinar e agir.” (Winston, 1992)</p>
Agindo como seres humanos	Agindo racionalmente
<p>“A arte de criar máquinas que executam funções que exigem inteligência quando executadas por pessoas.” (Kurzweil, 1990) “O estudo de como os computadores podem fazer tarefas que hoje são melhor desempenhadas pelas pessoas.” (Rich and Knight, 1991)</p>	<p>“Inteligência Computacional é o estudo do projeto de agentes inteligentes.” (Poole <i>et al.</i>, 1998) “AI... está relacionada a um desempenho inteligente de artefatos.” (Nilsson, 1998)</p>

Já Luger (2013, p.1) cita que “A Inteligência Artificial (IA) pode ser definida como o ramo da ciência da computação que se ocupa da automação do comportamento humano.”

5.1.1 Natural Language Processing (NLP)

Em 1950, Alan Turing lança o artigo “*COMPUTING MACHINERY AND INTELLIGENCE*” e propõe o atualmente conhecido Teste de Turing: uma pessoa conversa com um computador e com outro humano, sem saber qual deles é humano ou computador, e, se não for possível distinguir entre eles, então a máquina terá passado em seu teste. Por meio da técnica de *Natural Language Processing* (NLP), é possível à máquina compreender o que está sendo dito por um humano.

Para Sathiyakugan (2018), NLP é uma das disciplinas da IA conforme a Figura 1 apresenta. Trata-se da capacidade de um computador "entender" e gerar linguagem humana.

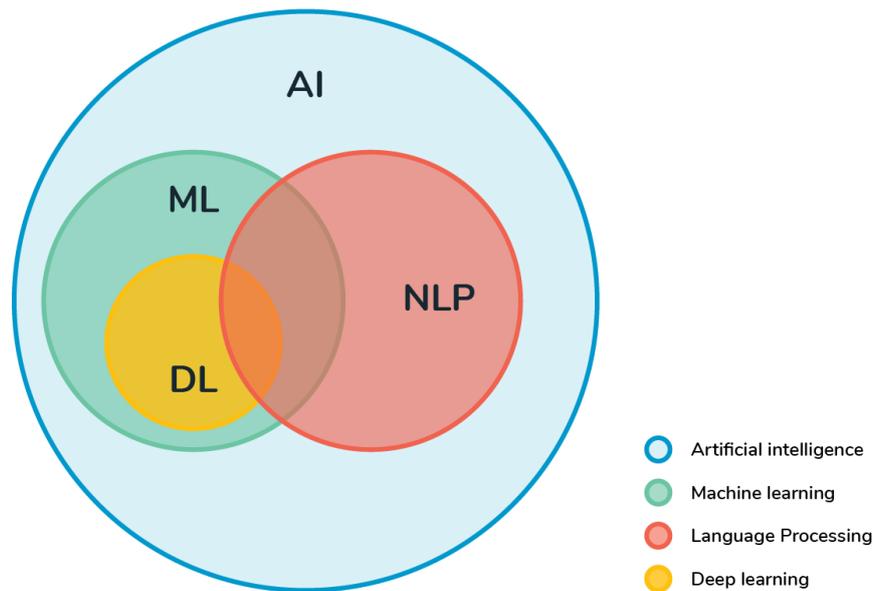


Figura 1: Subáreas de IA segundo Balakrishnan Sathiyakugan (2018, acessado em janeiro de 2020)

Luger (2013, p.515) lista diferentes níveis de análise para linguagem natural definidas por linguistas. São elas:

1. Prosódia, que trata do ritmo e entonação na linguagem;
2. Fonologia, responsável pelos sons que combinados formam a linguagem;
3. Morfologia, que se preocupa com a construção das palavras;
4. Sintaxe, onde são estudadas regras para construção de frases e sentenças válidas;
5. Semântica, responsável por estudar o significado transmitido em expressões em linguagem natural;
6. Pragmática, que trata do estudo das formas de uso da linguagem e de seus efeitos sobre o interlocutor;
7. Conhecimento do mundo, que inclui o conhecimento do mundo físico, social humano, com papéis, objetivos e intenções na comunicação.

Na maioria dos casos os algoritmos de NLP se preocupam principalmente em formalizar a sintaxe e representar a semântica e, dessa forma, tentam dar uma interpretação ao significado no mundo real. A Figura 2 exibe essa relação: há uma sequência entre os estágios na formação de uma sentença, passando pela análise sintática, interpretação semântica até o conhecimento contextual do mundo.

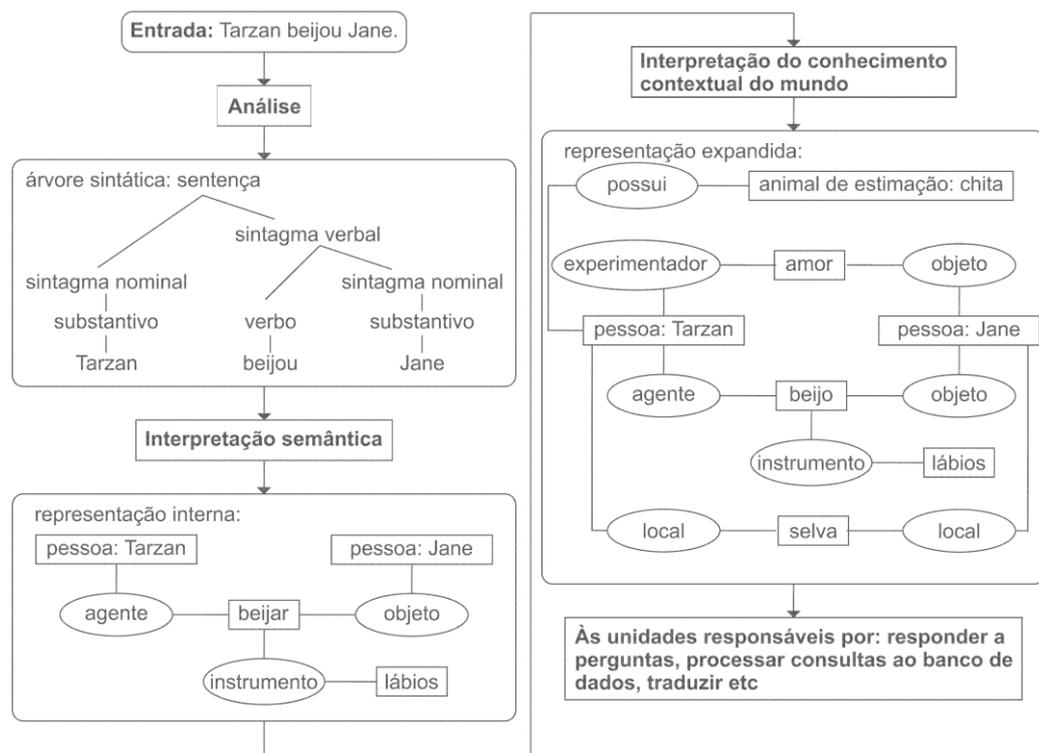


Figura 2: Estágios na produção de uma representação interna de uma sentença - Luger (2013, p.516)

É pela análise sintática que ocorre a verificação léxica comparando estruturas gramaticais definidas por cada língua. No entanto, neste momento não há identificação do significado da sentença, que fica a cargo da análise semântica.

NLP é utilizado em diferentes domínios, como recuperação da informação, mineração de textos, análises de sentimentos, traduções e transcrições de textos automáticos, Assistentes Virtuais, entre outros (EISENSTEIN Jacob, November, 2018). No próximo tópico, será abordado o conceito sobre Assistentes Virtuais e suas aplicações.

5.2 Assistentes Virtuais

Existe uma discussão sobre a diferença conceitual entre Assistentes Virtuais e *Chatbots*, na qual uma corrente trata o Assistente Virtual com capacidades adicionais ao *Chatbot*, como gerenciamento de contextos, execução de tarefas e escopo de atuação mais genérico. Em contrapartida há outra linha de pensamento que defende que ambos os termos são sinônimos e que classificam aplicações que têm como principal característica, a capacidade de interação de forma conversacional. Visto isso, o objetivo deste trabalho não é fomentar a discussão a respeito

da diferença ou não, e sim comparar as plataformas de desenvolvimento para *bots*. Para um entendimento comum, será utilizado *chatbots* seguindo a definição a seguir:

A definição clássica de um chatbot é um programa de computador que processa a entrada em linguagem natural de um usuário e gera respostas inteligentes e relativas que são então enviadas de volta ao usuário. Atualmente, os chatbots são movidos por mecanismos orientados por regras ou por dispositivos inteligentes (AI) que interagem com os usuários por meio de uma interface baseada em texto principalmente (KHAN; DAS, 2018, p. 1).

Porém existem algumas diferenças entre os tipos de *chatbots*, sobre as tecnologias e técnicas que utilizam, bem como seu escopo de atuação. Berger e Ebeners (2019) dividiram os tipos de chatbots conforme demonstrado na Figura 3.

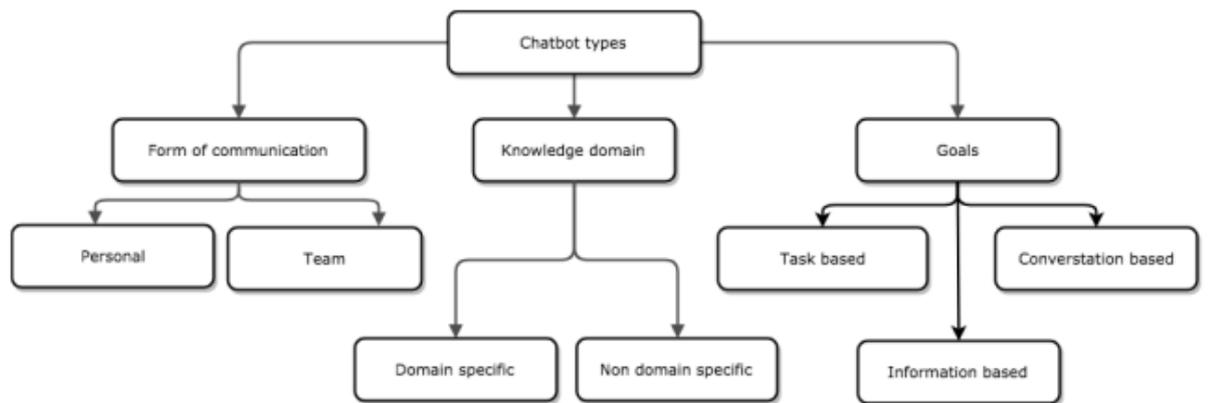


Figura 3: Tipos de *Chatbots* - BERGER e EBENERS (2019, Conception of a Conversational Interface to Provide a Guided Search of Study Related Data p.39)

Tipos de *Chatbots*:

- 1) Forma de Comunicar:
 - a) Pessoal: Um *bot* pessoal satisfaz as necessidades de um único usuário em um único contexto. Um exemplo típico de um *bot* pessoal é um assistente pessoal;
 - b) Em Equipe: O *bot* de equipe precisa alternar entre várias entradas do usuário e é usado em um canal compartilhado. Esses *bots* podem ser utilizados para organização da equipe em aplicativos de mensagens;
- 2) Domínio de conhecimento:
 - a) Domínio Específico: São geralmente implementados para um único serviço ou produto específico. Representam mais ou menos um produto ou uma marca;
 - b) Sem Domínio Específico: Expõe vários serviços, é o chamado *super bot*. Um *bot* muito conhecido dessa categoria é o Alexa da Amazon;

3) Objetivos:

- a) Baseado em Tarefas: Geralmente é implementado para executar uma determinada tarefa. O fluxo de conversação é predefinido e o objetivo principal é concluir um trabalho;
- b) Baseado em informações: Fornece informações para tópicos específicos. A conversa deve ser curta e objetiva. Um exemplo típico para esta categoria é um *bot* de perguntas frequentes (FAQ);
- c) Baseado em conversas: Esse tipo de *chatbot* tenta se comunicar com o usuário o maior tempo possível sem executar um trabalho específico. O principal objetivo dessa abordagem é uma conversa contínua com o usuário.

Nesse contexto existem várias plataformas de desenvolvimento de *Chatbots*, entre elas, Dialogflow do Google, Watson Assistant da IBM e Alexa Developer Console da Amazon. Em comum, todas estas plataformas utilizam Intenções como parte da compreensão sobre a interação do usuário. A IBM define intenção como:

Intenções são objetivos ou propósitos expressos em uma entrada do usuário, como a resposta a uma pergunta ou o processamento de um pagamento de fatura. Ao reconhecer a intenção expressa na entrada de um cliente, o serviço Watson Assistant pode escolher o fluxo de diálogo correto para responder a isso. (Site IBM, janeiro de 2021).

É na intenção onde existe a principal estrutura cognitiva de compreensão de linguagem natural, um exemplo é apresentado na Figura 4. Baseada em interações adicionados a cada intenção criada, a plataforma gera em seu modelo a capacidade de generalização sobre os exemplos adicionados de interações de usuários, como apresentado na Figura 5.

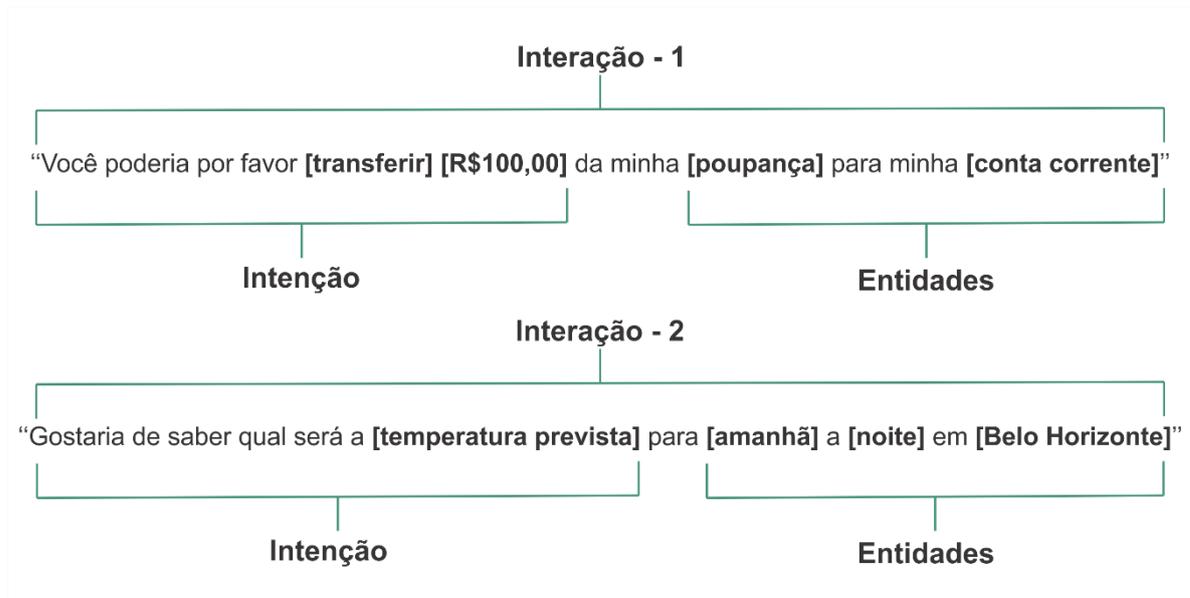


Figura 4: Intenção adaptada de *Adopting the power of conversational UX | Chatbots - Deloitte.co.uk/FS*

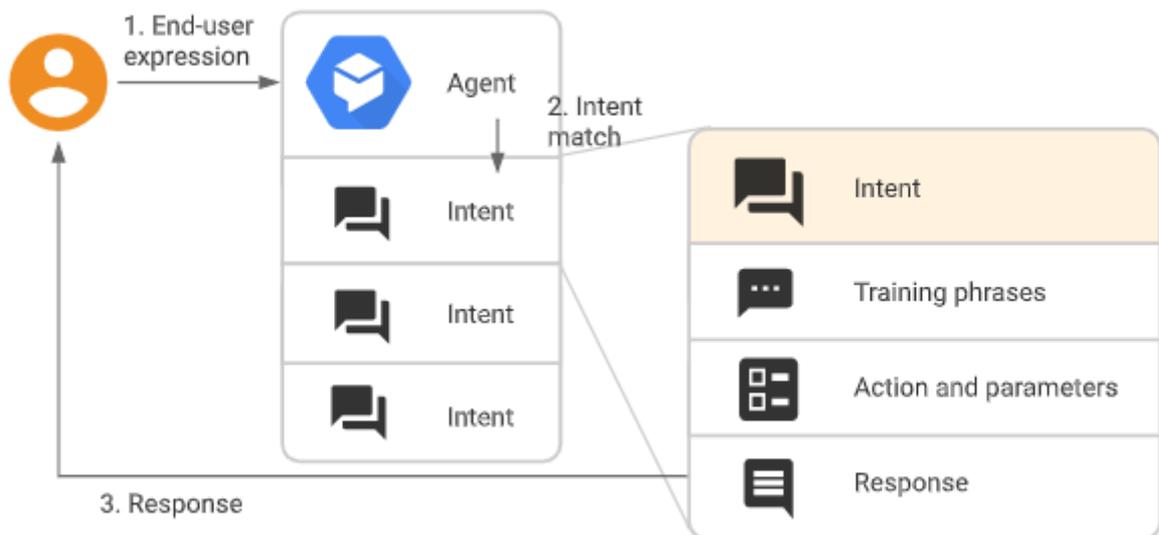


Figura 5: Diagrama do fluxo básico para correspondência de intenções e resposta ao usuário final (<https://cloud.google.com/dialogflow/docs/intents-overview>)

O principal objetivo deste trabalho é a comparação entre essas estruturas em diferentes plataformas, a fim de descobrir se existe ou não diferença entre as máquinas cognitivas de cada plataforma para as intenções em português do Brasil.

Em todas as plataformas é utilizado um *score* chamado confiança, que nada mais é que o percentual de certeza na classificação daquela interação, ou seja, quanto mais próximo de 1,

maior é a certeza do modelo que aquela classificação foi realizada corretamente, quanto mais próximo de 0, menor é a certeza da classificação correta.

6 METODOLOGIA

O experimento foi realizado considerando três fatores: Plataforma, Classe Intenção e Tipo de Intenção, com 3, 2 e 5 níveis cada, respectivamente, como ilustra a Figura 6.

6.1 Variável Resposta

Foi considerada como variável resposta qualitativa dicotômica, se o treinamento realizado na plataforma foi capaz de acertar ou não a classificação da interação do conjunto de teste, de acordo com o padrão ouro (*ground truth*) classificado pelo humano. Os dois resultados possíveis foram classificados em 1 (se acertou) e 0 (se não acertou) para as classificações diferentes do *ground truth* como primeira intenção classificada.

Apesar das plataformas considerarem a confiança para poder classificar a interação enviada do conjunto de teste, não foi tratado neste trabalho a diferença entre as confianças de cada classificação. Foi utilizada somente a variável binária (acertou ou não) com os três fatores (Plataforma, Classe Intenção e Intenção) como objeto desse experimento.

6.2 Criação de intenções

O processo de criação de uma intenção é simples. Dentro da intenção “Saber Previsão do Tempo”, por exemplo, pode-se ter as seguintes iterações, “Qual a temperatura agora?”, “Quantos graus está fazendo”, “Vai chover hoje”, “Como vai ser o dia hoje” entre outros. Basicamente deve-se adicionar exemplos reais de interações dos usuários a cada intenção criada, neste trabalho foram adicionados exemplos de interações textuais e não de fala. Essas interações são frases que caracterizam uma solicitação, uma resposta, ou mesmo um questionamento como apresentado na Figura 4.

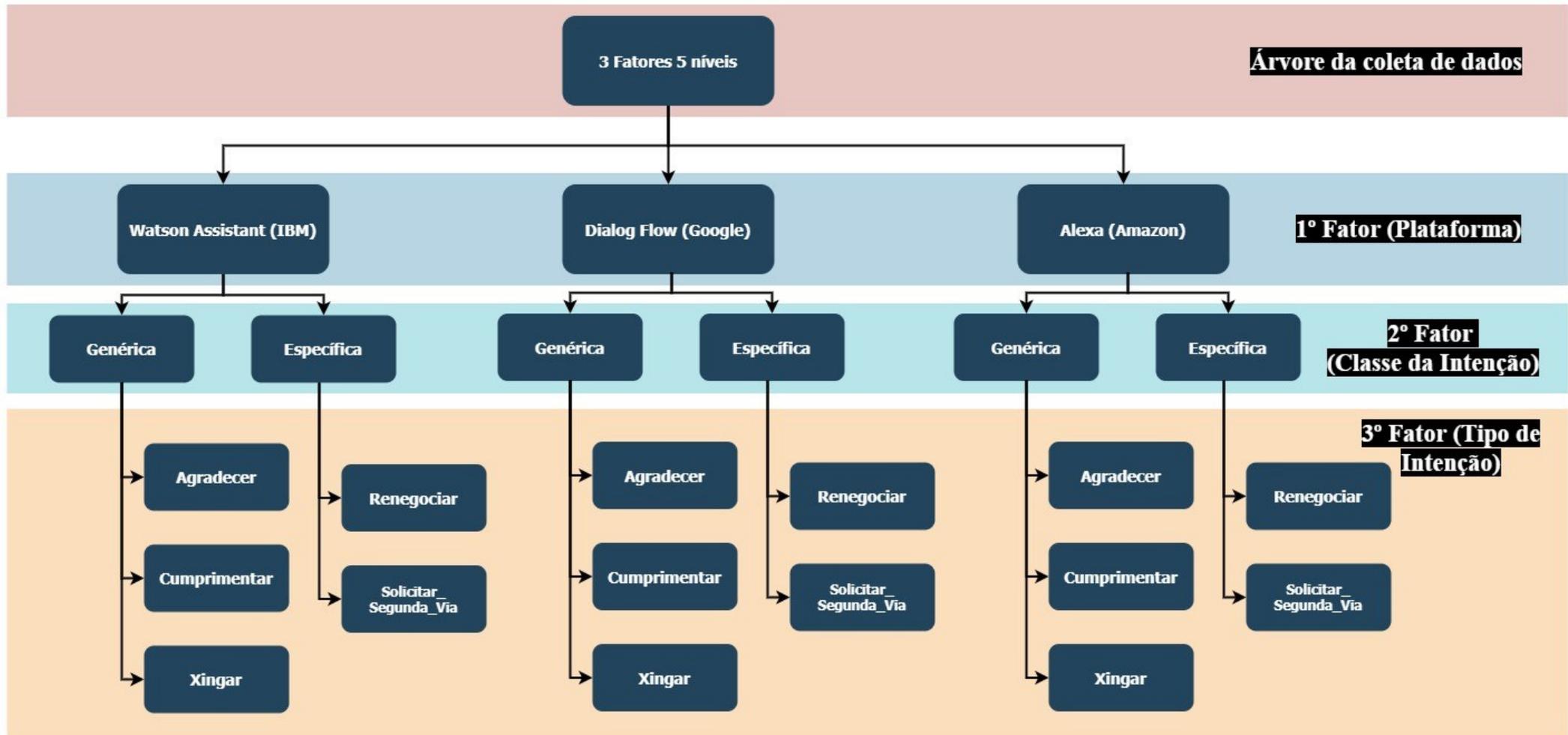


Figura 6: Matriz de definição das combinações entre fatores e níveis.

Para cada plataforma, existem especificações mínimas de configuração dessas intenções, por exemplo: no Watson Assistant da IBM a quantidade mínima de exemplos em uma intenção são cinco. Já para o Dialogflow não existe quantidade mínima, porém o indicado são 10 exemplos, e, para Alexa, também não existe quantidade mínima, apesar da documentação falar que devem ser adicionados a maior quantidade de exemplos possíveis sem explicitar uma quantidade.

Para esse experimento foi considerado como principal fator a **plataforma**, que possui três níveis, são eles:

Alexa Developer Console (Amazon) é a plataforma de desenvolvimento de assistentes virtuais criado pela Amazon e disponibilizado no Brasil em Abril de 2019, essa plataforma pode ser acessada pelo link: <https://developer.amazon.com/alexa/console/ask>.

Dialogflow (Google), inicialmente criado pela empresa Speaktio com o nome de api.ai em 2014, e adquirido pelo Google em setembro de 2016, sendo lançado com o nome de Dialogflow em outubro de 2017, é a plataforma do Google para o desenvolvimento de Assistentes Virtuais.

Watson Assistant (IBM) é a solução da IBM para o desenvolvimento de Assistentes virtuais e se encontra dentro da solução Watson lançada em 2011.

Um segundo fator avaliado neste trabalho foi a **classe da intenção** com dois níveis, genérico e específico, sendo o nível genérico composto por três intenções e o específico por duas intenções.

O tipo de intenção foi considerado o terceiro fator avaliado e contou com 5 níveis. Os exemplos de cada intenção e a descrição de cada nível está presente no esquema a seguir:

Genéricas: são intenções que podem ser utilizadas em qualquer contexto, independente do domínio de negócio existente. Para esse grupo foram criadas três intenções com 30 exemplos em cada, são elas:

- 1) Agradecer: alguns exemplos são, "obrigado por tudo", "ok obrigada", "Por ora não obrigado", "Que bom que Deus abençoe cada vez mais", "tenha um bom dia e obrigada", "tenha uma boa tarde valeu!", "Tudo bem obrigado", "continuar mais tarde mas te agradeço"...
- 2) Cumprimentar: alguns exemplos são, "como está essa força aí?", "como vai", "ei como vai?", "oi boa noite", "oi boa tarde", "Oi vc tá joia?", "olá ta tudo bem?", "Tudo bem"...
- 3) Xingar: alguns exemplos são, "atendimento é uma porcaria", "atendimento lixo", "Atendimento virtual é uma bosta", "Esta me tirando do sério", "Meu Deus devo está faltando grego", "nada que eu peço aqui sou atendida", "não gosto de você"...

Específicas: são intenções criadas em um domínio específico, por exemplo para o setor de comércio de bens duráveis. Dentro desse domínio existem vários contextos como, comercial, o financeiro, o de *marketing* e, nesse experimento, foi utilizado o contexto financeiro, com as intenções para a solicitação de segunda via de boleto e renegociar uma dívida. Cada uma das intenções teve 50 exemplos:

- 1) Solicitar_Segunda_Via: alguns exemplos são: “boletos de fevereiro para pagamento”, "código de barras do boleto vencimento", "conseguir outro boleto atualizado por aqui", "estou querendo imprimir um boleto tanto pelo site como pelo aplicativo", "estou querendo pagar a parcela do mes de maio"...
- 2) Renegociar: alguns exemplos são "com algumas parcelas em atraso simular uma renegociação do meu contrato", "com atraso em alguns pagamentos mais nao tenho a condição de pagar tudo no momento", "como faço para renegociar minhas dividas", "como fazer para parcelar minhas faturas vencidas", "dessas parcelas em atraso nao da para parcelar", "estar renegociando duas parcelas"...

Nota-se que, os cinco níveis do fator Tipo de Intenção puderam ser agrupados nos dois níveis do fator Classe de Intenção, Genérica e Específica. Sendo assim, para a análise dos dados do experimento, foram ajustados os modelos de regressão logística em separado para cada um dos dois fatores.

6.3 Criando o conjunto de treino e teste

O conjunto de treino foi utilizado para a criação de cada uma das intenções, como exposto na seção 6.2, esse conjunto teve origem em interações reais de usuários com chatbots, gerando uma base privada de interações.

Para cada um dos cinco tipos de intenções criadas, existe um conjunto de 30 interações, as quais não foram inseridas no conjunto de treinamento das intenções como exemplos. Todas essas interações foram classificadas manualmente por humanos conforme apresentado no quadro 2, sendo esse considerado o padrão ouro (*ground truth*), que será usado para avaliar a classificação de cada uma das plataformas.

Quadro 2: Iterações utilizadas para o teste e comparação dos modelos treinados em cada plataforma.

CONJUNTO DE TESTE			
Intenção (Nível)	Alguns Exemplos	Classe Intenção	Qtd. de interações
Agradecer	"muito obrigado pelo atendimento, fiquei super satisfeita" "Por ora não obrg" "gostei do atendimento, abraços" "tks, até mais..." "thanks! Conseguiu me ajudar.." "Fica com Deus, obrigadoooo!"	Genérica	30
Cumprimentar	"tudo certo por ai?" "opa blz" "hey boa tarde" "boa noitr, tudo certo por aí?" "me chamo daniel, tudo bem?" "fala comigo, blz"	Genérica	30
Xingar	"essa ineligência artificial não entende nada" "vc ta precisando de mais treinamento" "puta que pariu pra vc ser ruim tem que melhorar muito" "não me ajudou em nada, pessimo" "experiência pessima, como vcs têm coragem de atender com isso" "atendimento burro" "muito ruim credo"	Genérica	30
Solicitar_Segunda_Via	"boleto atualizado que venceu no dia" "boletos atualizados para pagamento" "emitir um novo boleto de pagamento" "você poderia me enviar este boleto" "realizar esse pagamento em atraso" "por favor quero minha segunda via"	Específica	30
Renegociar	"conversar sobre renegociação" "Quero fazer uma renegociação" "pagar minha entrada a vista" "Tenho uma dívida com vocês" "Tenho débitos, regularizar" "proposta de renegociação" "pagar meu debito com vocês"	Específica	30

6.4 Classificando em cada plataforma

Após o treinamento em cada plataforma com os cinco tipos de intenções, foi enviado para cada plataforma o conjunto de testes tratadas no item 6.2.

- 1) Para o Watson Assistant foi enviado via API com a seguinte chamada em linguagem Python exemplificada na Figura 7:

```

1  !pip install --upgrade "ibm-watson>=4.3.0"
2
3  from ibm_watson import AssistantV1
4  from ibm_watson import AssistantV2
5
6  assistant = AssistantV2(
7      |   version='2020-04-01',
8      |   authenticator= IAMAuthenticator('Seu_IAM')
9  )
10 response = assistant.message(
11     |   workspace_id='sua_skill',
12     |   alternate_intents = True,
13     |   input={
14     |       |   'text': 'mais tarde entro em contato de novo vlw'
15     |   }
16     |   ).get_result()
17

```

Figura 7: Exemplo de chamada da API Watson Assistant em Código Python, enviando um texto (iteração) e obtendo as intenções classificadas com as confianças de cada uma.

A Figura 8 apresenta o retorno em um Data Frame Pandas:

intent	Agradecer	Cumprimentar	Renegociar	Solicitar_Segunda_Via	Xingar	Interação
confidence	0.906248	0.294268	0.294274	0.246358	0.207569	mais tarde entro em contato de novo vlw
confidence	0.158938	0.170560	0.134657	0.119318	0.157520	Desculpas, apertei por engano. Obrig!
confidence	0.885864	0.312404	0.249172	0.244577	0.221599	vou continuar mais tarde, grata
confidence	0.998344	0.200114	0.223971	0.224215	0.260498	muito obrigado pelo atendimento, fiquei super ...
confidence	0.339781	0.274672	0.272964	0.264627	0.476418	Por ora não obrg

Figura 8: Retorno Watson Assistant IBM convertido em um Data Frame com iterações de teste classificadas em intenções com a confiança de cada uma.

- 2) Para o Dialogflow do Google também foi realizada uma chamada via api com a seguinte código em Python exemplificado na Figura 9:

```

1  !pip install dialogflow
2
3  import dialogflow_v2 as dialogflow
4  from google.colab import auth
5
6  df = pd.DataFrame()
7
8  texts = Validacao
9  project_id= DIALOGFLOW_PROJECT_ID
10 session_id= SESSION_ID
11 language_code= "pt-BR"
12
13 df = detect_intent_texts(project_id, session_id, texts, language_code)
14

```

Figura 9: Chamada a API do Dialogflow em Python para a classificação de iteração em uma intenção.

O retorno em um Data Frame Pandas é apresentado na Figura 10:

mais tarde entro em contato de novo vlw	Cumprimentar
Desculpas, apertei por engano. Obrg!	Cumprimentar
vou continuar mais tarde, grata	Cumprimentar
muito obrigado pelo atendimento, fiquei super ...	Cumprimentar
Por ora não obrg	Cumprimentar

Figura 10: Retorno Dialogflow com iterações de teste classificadas em uma intenção.

- 3) Para a Alexa Amazon existiu uma particularidade nos dados de treinamento e teste, pois a plataforma não aceita caracteres especiais como “, - ! .”. Em ambos os conjuntos, foram retirados os caracteres especiais e a chamada foi realizada de maneira manual de cada iteração de teste como apresentado na Figura 11, e a saída adicionada em uma planilha:

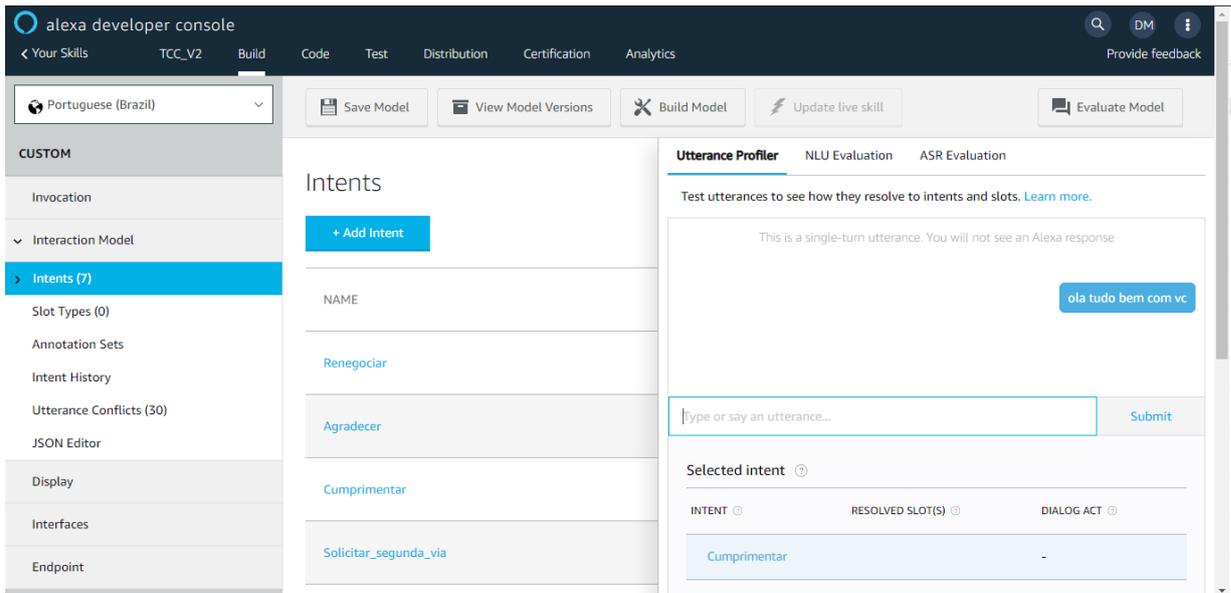


Figura 11: Console Alexa Developer (<https://developer.amazon.com/alexa/console/ask>)

6.5 Hipóteses de trabalho

Hipótese Nula Principal: não existem diferenças dos motores cognitivos das plataformas Dialogflow, Watson Assistant e Alexa Skill Kit quanto ao acerto nas classificações das intenções.

Hipótese Nula Secundária: não existem diferenças da classe ou do tipo de intenção quanto ao acerto nas classificações das intenções.

6.6 Regressão logística

Para estudar o comportamento e diferença entre os fatores e níveis, foi utilizado a regressão logística.

A característica da variável resposta Y_i também chamada de dependente, é ser binária, ou seja, somente são possíveis dois valores, sucesso para os casos em que a resposta assume valor igual a 1 e fracasso para respostas não corretas, que assumem valor 0. Desse modo, Y_i binária segue uma distribuição de *Bernoulli* em que $P(Y_i = 1) = \pi_i$ é a probabilidade de sucesso e $P(Y_i = 0) = 1 - \pi_i$ é a probabilidade de fracasso. O valor esperado de Y_i é $E(Y_i) = \pi_i$ e variância igual a $\text{Var}(Y_i) = \pi_i (1 - \pi_i)$.

No modelo de regressão logística binária, a relação entre a probabilidade de sucesso, $E(Y)$, e as variáveis independentes é modelada por meio da função *logit* $f(z) = \ln(z/(1-z))$, com $-\infty < z < \infty$. Assim, para um modelo de regressão logística com uma variável independente x , podemos escrever $E(Y) = \frac{\exp(\beta_0 + \beta_1 x)}{1 + \exp(\beta_0 + \beta_1 x)}$. O comportamento da função logit é ilustrado na figura 12, que

mostra que o valor de $E(Y)$, a probabilidade de sucesso, fica confinado ao intervalo de entre 0 e 1, qualquer que seja o valor de x

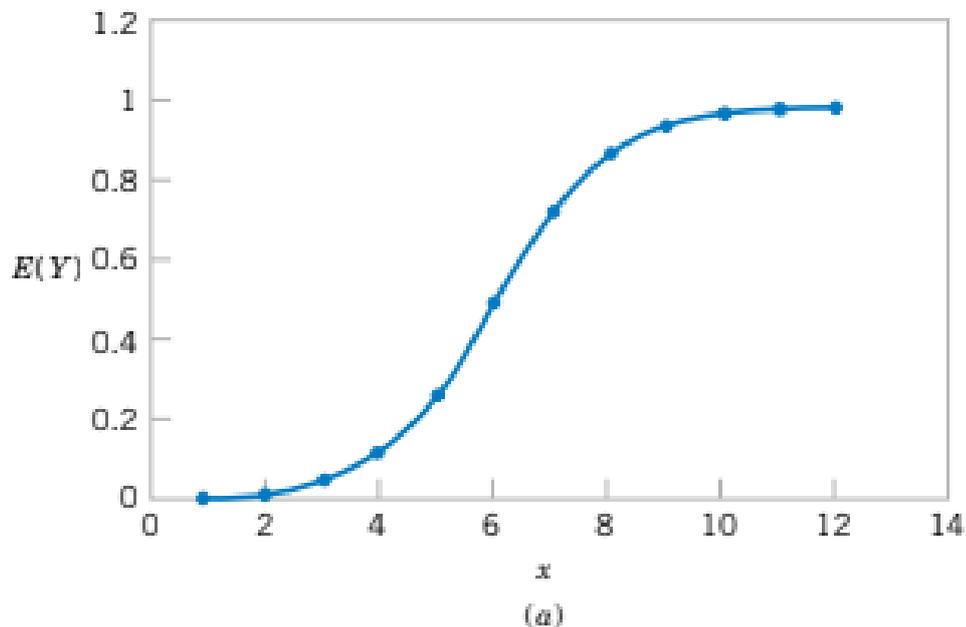


Figura 12: Exemplo da função logística para o caso em que $E(Y) = 1 / (1 + e^{-6 \cdot x})$ - MONTGOMERY e RUNGER (2012)

Odds em sua forma mais simples, é a razão entre a probabilidade de um evento acontecer e a probabilidade do evento não acontecer (DG Kleinbaum, M Klein, 2010). Por exemplo, suponhamos que joguemos uma moeda 100 vezes e, em 50 dessas vezes, o evento coroa aconteça e nas outras 50 o evento coroa não aconteça, pois deu cara. A odds de coroa é de $50/50 = 1$

$$Odds = \frac{P}{1-P}$$

$$Odds: P(X) = \frac{1}{1+e^{-(x+\sum \beta_i X_i)}}$$

Já *odds ratio* é a medida de associação diretamente estimada a partir de um modelo logístico sem premissas especiais, e ainda, qualquer *odds ratio*, por definição, é uma razão de duas *odds* escrito como $odds_1$ dividido por $odds_0$, onde os índices indicam dois grupos ou indivíduos diferentes sendo comparados (DG Kleinbaum, M Klein, 2010).

$$Odds Ratio (OR) = \frac{odds_1}{odds_0}$$

Por exemplo, supondo que existam dois grupos de cartas, G1 e G2. No grupo G1 a *odds* de ganhar é de 0,67, ou seja, a probabilidade de ganhar é de 0,40 e a de perder é 0,60. Já no grupo G2 a *odds* de ganhar é de 0,49, sendo a probabilidade de ganhar 0,33 e a de perder 0,67%. Em

ambos os casos a probabilidade de perder é maior, por isso a *odds* de ganhar para os dois grupos é menor do que 1. No entanto, a *odds ratio* de ganhar comparando o G1 com o G2 é de 1,36, ou seja, 0,67/0,49. A odds de ganhar com o grupo de cartas G1 é 36% maior do que a odds de ganhar com o grupo de cartas G2.

6.7 Modelo de regressão logística para os fatores plataformas e classe de intenção

Para a análise multivariada dos fatores Plataforma e Classe de Intenção, foi utilizado o seguinte modelo:

$$\ell^n\left(\frac{p}{1-p}\right) = \beta_0 + \beta_{11}C_1 + \beta_{12}C_2 + \beta_{21}CI_1$$

Onde, p é a probabilidade de acerto da classificação

$$C_1 = \begin{cases} 1, & \text{se Alexa} \\ 0, & \text{c.c.} \end{cases}$$

$$C_2 = \begin{cases} 1, & \text{se IBM Watson} \\ 0, & \text{c.c.} \end{cases}$$

$$CI_1 = \begin{cases} 1, & \text{se Genérica} \\ 0, & \text{c.c.} \end{cases}$$

Sendo utilizado o Dialogflow como classe de referência para o fator plataforma e Específica como classe referência para Classe Intenção.

6.8 Modelo de regressão logística para os fatores plataforma e tipo de intenção

Para a análise multivariada dos fatores Plataforma e Tipo de Intenção, foi utilizado o seguinte modelo:

$$\ell^n\left(\frac{p}{1-p}\right) = \beta_0 + \beta_{11}C_1 + \beta_{12}C_2 + \beta_{21}I_1 + \beta_{22}I_2 + \beta_{23}I_3 + \beta_{24}I_4$$

Onde, p é a probabilidade de acerto da classificação, C1 e C2 são definidos como na seção 6.7 e

$$I_1 = \begin{cases} 1, & \text{se Agradecer} \\ 0, & \text{c.c.} \end{cases}$$

$$I_2 = \begin{cases} 1, & \text{se Cumprimentar} \\ 0, & \text{c.c.} \end{cases}$$

$$I_3 = \begin{cases} 1, & \text{se Xingar} \\ 0, & \text{c.c.} \end{cases}$$

$$I_4 = \begin{cases} 1, & \text{se Renegociar} \\ 0, & \text{c.c.} \end{cases}$$

Também utilizado o Dialogflow como classe de referência para o fator plataforma e Segunda Via como classe referência para Tipo de Intenção

Para gerar o modelo Modelo Binário de Regressão Logística e as comparações foi utilizado o Software Minitab versão 17.

7 RESULTADOS

7.1 Análise descritiva dos dados

A Tabela 1 apresenta a análise descritiva dos resultados do experimento.

Tabela 1 – Percentual de acertos segundo cada tratamento utilizado no experimento (tipo de plataforma x tipo de interação)

Segundo Fator Classe da Intenção	Terceiro Fator Tipo de Intenção	Primeiro Fator (Plataforma)			Média
		Watson Assistant (IBM)	Alexa Developer (Amazon)	Dialogflow (Google)	
Genérica	Agradecer	80,0%	43,3%	0,0%	41,1%
Genérica	Cumprimentar	76,7%	43,3%	23,3%	47,8%
Genérica	Xingar	80,0%	63,3%	56,7%	66,7%
Média do percentual de acerto (Genérica)		78,9%	50,0%	26,7%	51,9%
Específica	Renegociar	90,0%	73,3%	93,3%	85,6%
Específica	Solicitar_Segunda_Via	83,3%	76,7%	80,0%	80,0%
Média do percentual de acerto (Específica)		86,667%	75,0%	86,667%	82,8%
Média do percentual total de acertos		82,0%	60,0%	50,7%	64,2%

Na imagem 13, é possível perceber de forma visual a diferença nos percentuais de acertos entre as classes de intenções e plataformas:

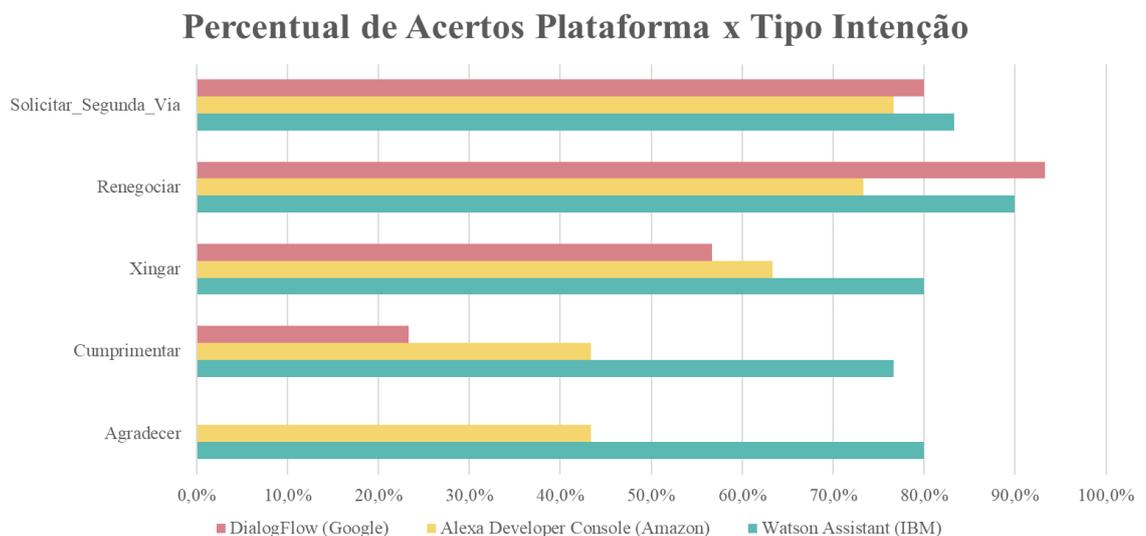


Figura 13 Gráfico de percentual de acertos.

A plataforma Watson Assistant IBM obteve um percentual médio de acerto de 82%, ou seja, das 150 iterações enviadas para teste, 123 tiveram classificações da primeira intenção (com maior confiança) iguais da intenção classificada por humano, conforme percentual apresentado na tabela 1.

A plataforma Dialogflow do Google obteve um percentual de acertos de 50,7%, ou seja, das 150 iterações enviadas de teste, apenas 76 tiveram classificações da primeira intenção iguais à intenção classificada por humano.

A plataforma Alexa da Amazon obteve um percentual de acertos de 60%, ou seja, das 150 iterações enviadas para avaliação, 90 tiveram classificações da primeira intenção iguais à intenção classificada por humano.

Entre os níveis do fator Classe de Intenção, é possível perceber que, no nível Específica, o percentual de acerto é cerca de 30% maior do que no nível Genérica.

E entre os níveis do fator Tipos de Intenções, a que teve maior percentual de acerto foi a Renegociar com uma média de 85,6%, sendo o tipo com a pior performance a Agradecer com uma média de acerto de 41,1%.

A Figura 14 apresenta o gráfico de efeitos principais dos três fatores. A análise desses gráficos indica possíveis diferenças na proporção de acertos de classificações entre os níveis de todos os três fatores.

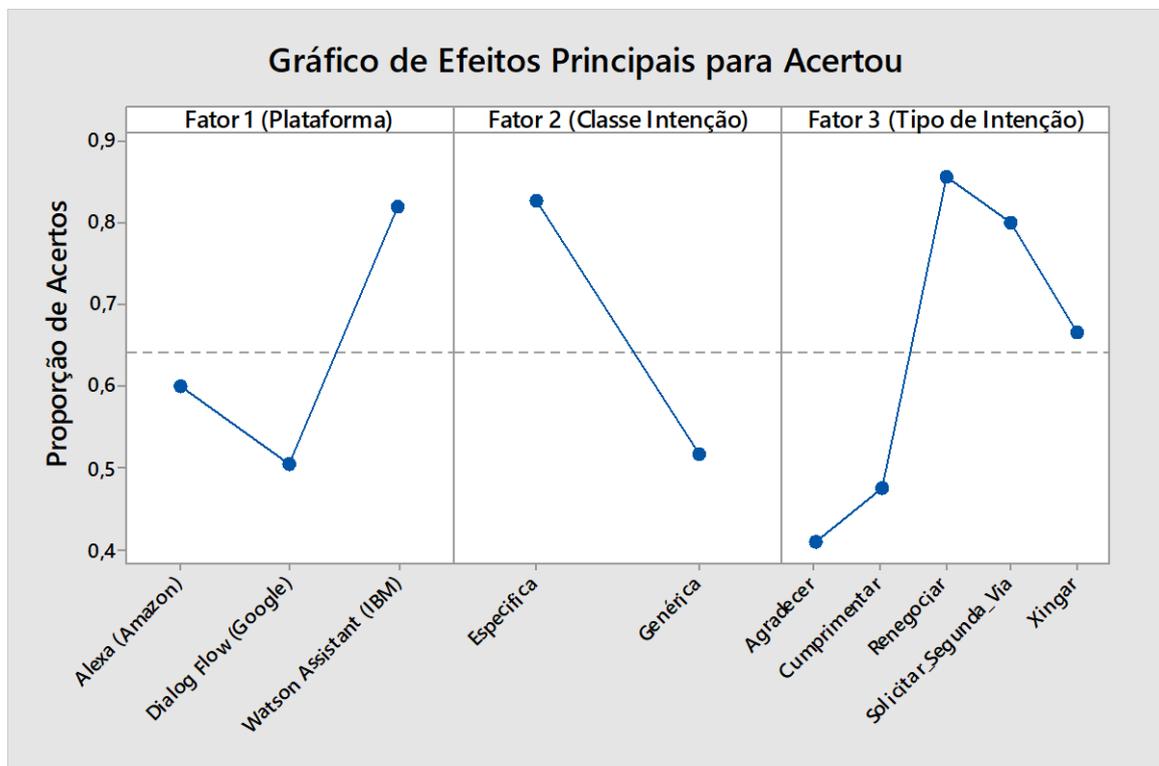


Figura 14 Gráfico de Efeitos Principais.

Na figura 15, é apresentado o gráfico de interação entre os fatores Plataforma e Tipo de Intenção, que mostra que a possibilidade da existência de interação entre esses dois fatores deve ser cuidadosamente investigada.

A figura 16 apresenta o gráfico de interação entre os fatores Plataforma e Classe de Intenção. Esse gráfico mostra que a possibilidade da existência de interação entre esses dois fatores também deve ser cuidadosamente investigada.

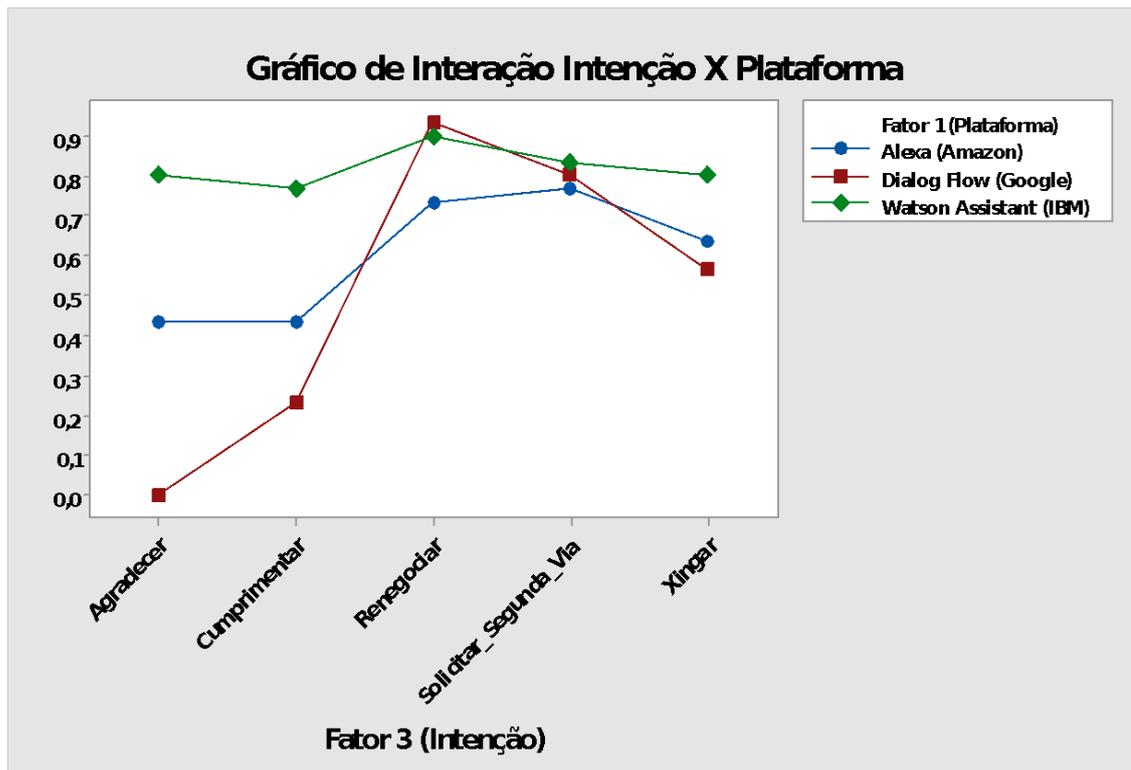


Figura 15: Gráfico de interação entre os fatores Plataforma e Tipo de Intenção.

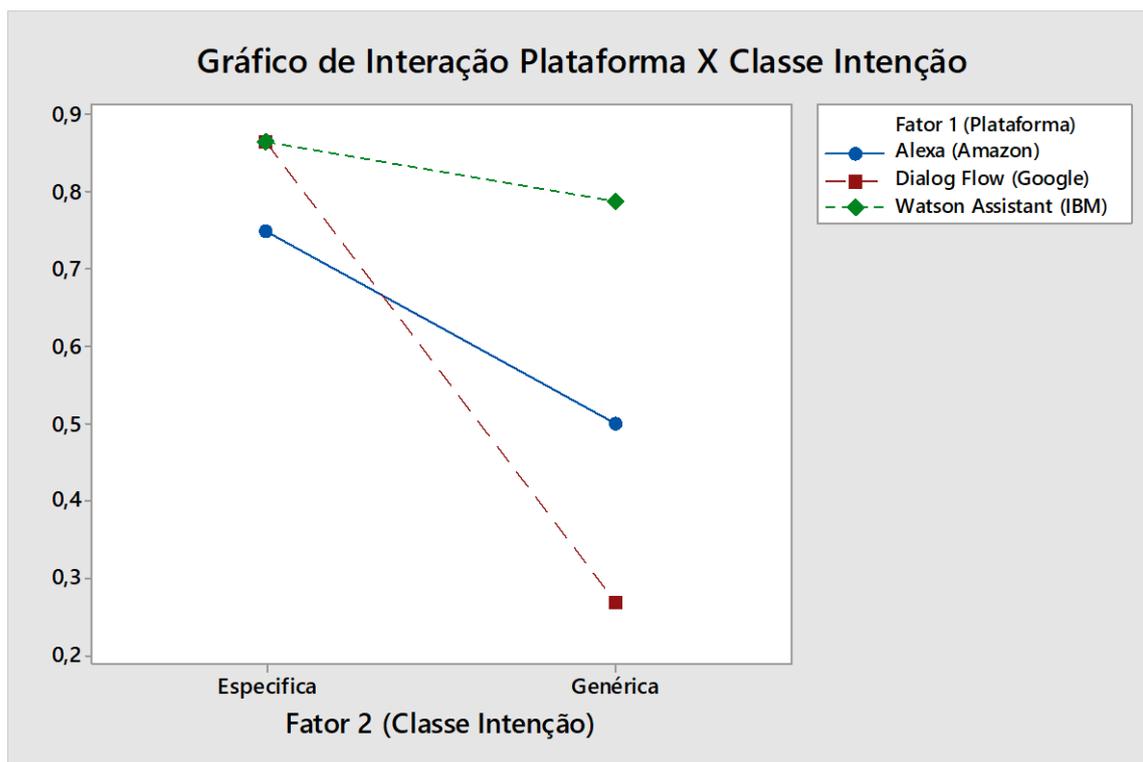


Figura 16: Gráfico de interação entre os fatores Plataforma e Classe da Intenção.

7.2 Resultados para cada um dos fatores

Na tabela 2 é apresentado um compilado das *odds ratios* de acerto para cada um dos níveis de fator separadamente (modelo simples).

Tabela 2 – Odds Ratio de acerto entre cada um dos níveis de cada fator separadamente com seus respectivos intervalos de confiança

Fator	Nível 1	Nível 2	Odds Ratio	Intervalo de Confiança 95%	Significant e a 5%
Fator 1 (Plataforma)	Dialogflow (Google)	Alexa (Amazon)	0,68	(0,43; 1,08)	Não
Fator 1 (Plataforma)	Watson Assistant (IBM)	Alexa (Amazon)	3,03	(1,78; 5,15)	Sim
Fator 1 (Plataforma)	Watson Assistant (IBM)	Dialogflow (Google)	4,43	(2,62; 7,50)	Sim
Fator 2 (Classe Intenção)	Específica	Genérica	4,46	(2,83; 7,03)	Sim
Fator 3 (Tipo de Intenção)	Cumprimentar	Agradecer	1,31	(0,72; 2,36)	Não
Fator 3 (Tipo de Intenção)	Renegociar	Agradecer	8,48	(4,12; 17,47)	Sim
Fator 3 (Tipo de Intenção)	Solicitar_Segunda_Via	Agradecer	5,73	(2,94; 11,15)	Sim
Fator 3 (Tipo de Intenção)	Xingar	Agradecer	2,86	(1,56; 5,26)	Sim
Fator 3 (Tipo de Intenção)	Renegociar	Cumprimentar	6,47	(3,16; 13,28)	Sim

Fator 3 (Tipo de Intenção)	Solicitar_Segunda_Via	Cumprimentar	4,37	(2,26; 8,47)	Sim
Fator 3 (Tipo de Intenção)	Xingar	Cumprimentar	2,19	(1,20; 3,99)	Sim
Fator 3 (Tipo de Intenção)	Solicitar_Segunda_Via	Renegociar	0,68	(0,31; 1,47)	Não
Fator 3 (Tipo de Intenção)	Xingar	Renegociar	0,34	(0,16; 0,70)	Sim
Fator 3 (Tipo de Intenção)	Xingar	Solicitar_Segunda_Via	0,50	(0,25; 0,98)	Sim

Como pode-se observar na Tabela 2, existem as seguintes *odds* entre as plataformas:

- 1) A Plataforma Watson da IBM tem uma *odds* de acertar a resposta 3 vezes maior do que a *odds* de acerto da Amazon;
- 2) Essa razão é ainda maior comparando com a Plataforma do Google. Nesse caso, a solução da IBM tem uma *odds* 4,43 vezes maior de acertar a resposta.
- 3) No entanto quando comparada as plataformas Dialogflow e Alexa, pode-se perceber que não há significância, visto que seu intervalo de confiança (IC) passa pelo 1.

A *odds* entre os níveis do fator 2 foram relevantes, uma vez que teve uma *Odds ratio* 4,46 que significa uma *odds* de acerto maior quase 4,5 vezes nos casos das intenções específicas, conforme apresentado na Tabela 2.

Em quase todos os testes realizados entre as intenções, foi possível perceber que o intervalo de confiança (IC) não passa pelo valor 1, o que torna a razão de *odds* (OR) estatisticamente significativa. Somente nos casos entre as intenções Cumprimentar e Agradecer, e entre Solicitar Segunda Via e Renegociar, não houve significância estatística. Os valores de *odds* mais significativos apresentados na Tabela 2 foram:

- 1) A maior diferença foi encontrada entre as intenções Renegociar e Agradecer, sendo que a *odds* de acerto da intenção Renegociar foi 8,48 vezes maior do que para a intenção Agradecer;
- 2) A segunda maior diferença foi encontrada entre as intenções Renegociar e Cumprimentar, sendo que a *odds* de acerto da intenção Renegociar foi 6,47 vezes maior do que a intenção Cumprimentar;
- 3) Com a razão de *odds* de 2,19, a intenção Xingar têm duas vezes mais *odds* de acerto do que intenção Cumprimentar
- 4) Em comparação a intenção Agradecer, a intenção Solicitar Segunda Via tem a *odds* de acerto 5,7 vezes maior;
- 5) No caso do teste entre as intenções Xingar e Agradecer, a intenção Xingar têm uma *odds* de acerto 2,86 vezes maior;

- 6) Entre Solicitar Segunda Via e Cumprimentar, Solicitar Segunda Via têm uma *odds* de acerto 4,3 vezes maior;
- 7) Na comparação entre Xingar e Renegociar, Renegociar têm uma 0,67 maior;
- 8) E entre Xingar e Solicitar Segunda Via, Solicitar Segunda Via têm uma OR de 0,5 maior.

7.3 Análise múltipla

Após analisar cada um dos fatores individualmente, foi gerado também o modelo múltiplo com a interação entre os fatores.

7.3.1 Modelo múltiplo – Interação entre classe de intenção e plataforma

A figura 17 traz os resultados do ajuste do modelo com interação entre os fatores Plataforma e Classe de Intenção. Quando adicionado ao modelo a interação entre os fatores Plataforma e Classe da Intenção, essa interação se mostra tão forte que o fator plataforma deixa de ser estatisticamente significativa ao nível de significância de 5%.

Source	DF	Adj Dev	Adj Mean	Chi-Square	P-Value
Regression	5	103,266	20,653	103,27	0,000
Fator 1 (Plataforma)	2	3,659	1,830	3,66	0,160
Fator 2 (Classe Intenção)	1	9,657	9,657	9,66	0,002
Fator 1 (Plataforma)*Fator 2 (Classe Intenção)	2	16,033	8,016	16,03	0,000
Error	444	483,650	1,089		
Total	449	586,916			

Figura 17: Resultado do modelo considerando fatores Plataforma, Classe de Intenção e interação entre eles.

7.3.2 Modelo simples – Segregado pelo nível genérico para o fator classe de intenção

Diante da evidência de significância estatística da interação entre os fatores Plataforma e Classe de Intenção, a comparação das plataformas foi feita em cada um dos dois níveis da Classe de Intenção. Assim, o conjunto de dados foi segregados em dois subconjuntos: o primeiro somente com os casos da Classe de Intenção Genérica, e o segundo, somente com o nível Específico para o mesmo fator.

A tabela 3 mostra os resultados da comparação entre plataformas na Classe de Intenção Genérica.

Tabela 3: Resultado do modelo simples com o fator Plataforma, segregado pelo nível Genérico do fator Classe Intenção.

Níveis Fator Classe Intenção	Nível 1	Nível 2	Odds Ratio	Intervalo de Confiança 95%	Significant e a 5%
1 (Genérica)	Dialogflow (Google)	Alexa (Amazon)	0,36	(0,19; 0,68)	Sim
1 (Genérica)	Watson Assistant (IBM)	Alexa (Amazon)	3,74	(1,94; 7,18)	Sim
1 (Genérica)	Watson Assistant (IBM)	Dialogflow (Google)	10,28	(5,16; 20,46)	Sim

- 1) A plataforma Dialogflow (Google) tem uma *odds ratio* de acertar a resposta 0,36 menor do que a *odds* de acerto da Alexa (Amazon);
- 2) Para a plataforma Watson Assistant (IBM) há uma *odds ratio* 3,74 vezes maior de acerto do que a plataforma Alexa (Amazon)
- 3) Mas a maior disparidade está entre a plataforma Watson (IBM) e o Dialogflow (Google), onde a primeira apresenta uma *odds ratio* maior 10,287 vezes para acertar do que a segunda.

7.3.3 Modelo simples – Segregado pelo nível específico para o fator classe intenção

Quando ajustado o modelo somente com o nível Específico do fator Classe Intenção (Tabela 4). é possível ver que não há evidência de diferença entre as plataformas na classe de intenção específica, a nível de significância de 5%.

Tabela 4: Resultado do modelo simples com o fator Plataforma, segregado pelo nível Específico do fator Classe Intenção.

Níveis Fator Classe Intenção	Nível 1	Nível 2	Odds Ratio	Intervalo de Confiança 95%	Significant e a 5%
1 (Específica)	Dialogflow (Google)	Alexa (Amazon)	2,17	(0,84; 5,58)	Não
1 (Específica)	Watson Assistant (IBM)	Alexa (Amazon)	2,17	(0,84; 5,58)	Não
1 (Específica)	Watson Assistant (IBM)	Dialogflow (Google)	1,00	(0,34; 2,86)	Não

7.3.4 Modelo simples – Comparação entre plataformas para cada nível do fator tipo de intenção

O gráfico de interação entre os fatores Plataforma e Tipo de Intenção (Figura 14) mostra evidências de interação entre esses dois fatores para explicar a probabilidade de acerto da classificação. Assim, uma comparação entre as plataformas foi feita em cada um dos cinco níveis

do fator Tipo de Intenção. Na tabela 5 é apresentado um compilado entre todos os resultados dos modelos ajustados.

Tabela 5 – Odds Ratio entre cada um dos níveis de do fator Plataforma, segundo o Tipo de Intenção, e seus respectivos intervalos de confiança.

Classe Intenção	Níveis Fator Tipo Intenção	Nível 1	Nível 2	Odds Ratio	Intervalo de Confiança 95%	Significante
Genérica	Agradecer	Dialogflow (Google)	Alexa (Amazon)	-	-	-
Genérica	Agradecer	Watson Assistant (IBM)	Alexa (Amazon)	5,23	(1,65;16,51)	Sim
Genérica	Agradecer	Watson Assistant (IBM)	Dialogflow (Google)	-	-	-
Genérica	Cumprimentar	Dialogflow (Google)	Alexa (Amazon)	0,40	(0,13; 1,21)	Não
Genérica	Cumprimentar	Watson Assistant (IBM)	Alexa (Amazon)	4,29	(1,41;13,06)	Sim
Genérica	Cumprimentar	Watson Assistant (IBM)	Dialogflow (Google)	10,80	(3,26;35,71)	Sim
Genérica	Xingar	Dialogflow (Google)	Alexa (Amazon)	0,76	(0,26;2,13)	Não
Genérica	Xingar	Watson Assistant (IBM)	Alexa (Amazon)	2,32	(0,72;7,40)	Não
Genérica	Xingar	Watson Assistant (IBM)	Dialogflow (Google)	3,06	(0,96;9,65)	Não
Específica	Solicitar Segunda Via	Dialogflow (Google)	Alexa (Amazon)	1,22	(0,35;4,17)	Não
Específica	Solicitar Segunda Via	Watson Assistant (IBM)	Alexa (Amazon)	1,52	(0,42;5,47)	Não
Específica	Solicitar Segunda Via	Watson Assistant (IBM)	Dialogflow (Google)	1,25	(0,33;4,64)	Não
Específica	Renegociar	Dialogflow (Google)	Alexa (Amazon)	5,09	(0,98;26,42)	Não
Específica	Renegociar	Watson Assistant (IBM)	Alexa (Amazon)	3,27	(0,77;13,83)	Não
Específica	Renegociar	Watson Assistant (IBM)	Dialogflow (Google)	0,64	(0,09;4,15)	Não

*O nível Agradecer no Fator Tipo de Intenção foi realizado somente entre Alexa e Watson Assistant, como não houve nenhum acerto do Dialogflow, não foi possível ajustar o modelo.

O modelo com o *data set* segregado pelo nível Agradecer do fator Tipo de Intenção não pode ser ajustado, pois o nível Dialogflow do fator Plataforma não classificou nenhuma das 30 interações de testes na categoria correta. No entanto o modelo foi ajustado somente com os dois níveis Alexa e Watson Assistant no fator 1 plataforma, uma vez que o evento não ocorreu no nível Dialogflow, no fator classe da intenção somente com o nível agradecer. Nesse caso pode-se perceber uma *odds ratio* de 5,23 vezes maior de acerto para a plataforma Watson Assistant comparado a Alexa.

O modelo ajustado pelo *dataset* considerando somente o nível Cumprimentar no fator Tipo de Intenção apresentou significância somente entre as plataformas Watson Assistant (IBM) e Alexa (Amazon), sendo a *odds ratio* maior 4,29 vezes para a classificação correta na plataforma Watson Assistant, e em comparação com a plataforma Dialogflow (Google) a plataforma da IBM obteve uma *odds ratio* de 10,79 vezes maior.

Não houve significância estatística ao nível de 5% de confiança para nenhuma das comparações entre as plataformas, considerando apenas os *data sets* dos níveis Xingar, Solicitar Segunda Via e Renegociar no fator Tipo de Intenção.

Para nenhum dos modelos ajustados, o teste de bondade de ajuste de Hosmer-Lemeshow mostrou evidência de falta de ajuste dos modelos (valores- $p > 0,05$).

8 CONCLUSÕES

Com o principal objetivo de comparar a capacidade de generalização dos modelos cognitivos utilizados nas principais plataformas de desenvolvimentos de assistentes virtuais e *chatbots* (Alexa Console Developer da Amazon, Dialogflow do Google e Watson Assistant da IBM), esse trabalho tenta apoiar organizações que pretendem iniciar a construção de seus assistentes virtuais, na redução da complexidade da avaliação de qual plataforma escolher. Desse modo, essas organizações podem concentrar-se em outros fatores como custo, expertise, facilidade na utilização, estabilidade e maturidade da plataforma, facilidade nas integrações, logs, *roadmap* de evolução e suporte entre outros.

Para isso, foi planejado um experimento para o qual foram criadas interações de cinco tipos de intenções em cada uma das três plataformas avaliadas. Dessas cinco intenções, duas eram de um grupo considerado específicas (dentro de um contexto de negócio), as outras três consideradas genéricas (comuns em qualquer contexto de negócio). Para cada intenção do grupo genérico foram adicionados trinta exemplos de interações. Já para grupo específico, foram adicionados cinquenta exemplos de interações em cada. Essas interações compuseram o conjunto de treinamento das plataformas.

Como conjunto de teste das plataformas, foram geradas mais trinta novas interações para cada tipo de intenção.

Após rodar cada um dos trinta exemplos das cinco intenções em cada uma das três plataformas, o resultado foi compilado de forma binária, considerando apenas a classificação da primeira intenção de acordo com o padrão ouro (*ground truth*) criado anteriormente para cada uma das interações de testes, onde o valor 1 correspondia a classificação correta na primeira intenção e 0 quanto a classificação realizada pela plataforma era diferente do padrão ouro.

Com esse resultado foram ajustados modelos de regressões logística com o objetivo de verificar se havia ou não diferenças estatisticamente significantes entre as plataformas por meio do cálculo das *odds* de acerto da classificação, que foram comparadas via *odds ratio*.

A partir das análises dos dados do experimento, pode-se concluir que:

- 1) De forma geral, há diferença estatística significativa, ao nível de significância de 5%, entre as plataformas;
- 2) A plataforma que obteve melhor resultado foi a do Watson Assistant da IBM, com uma média global de acerto 82% e uma *odds ratio* de acerto geral de 3,03 comparada a Alexa e de 4,43 comparada ao Dialogflow.

3) Entre o Dialogflow e a Alexa, não foi encontrada significância estatística para a diferença geral, apesar da Alexa ter ficado com um percentual geral de acerto de 60% contra 50,7% do Dialogflow.

Quando segregamos os percentuais de acertos por grupo de classes da intenção (Genérica ou Específica), percebemos que a diferença de acerto percentual aumenta para o grupo Genérica, sendo que o Watson Assistant possui o maior percentual de acerto (78,9%), contra 50% da plataforma Alexa e 26,7% do Dialogflow. Para a classe de intenção Genérica, a *odds ratio* de acerto comparando Watson Assistant e Dialogflow é de 10,28 (IC 95% 5,16; 20,46), e de 3,74 quando se compara Watson Assistant e Alexa (IC 95% 1,94; 7,18). Entre Dialogflow e Alexa, não houve diferença estatística entre as odds de acerto na classe Genérica.

Na classe específica, também há uma diferença no comportamento de acertos das três plataformas, sendo a estimativa pontual do percentual de acerto médio igual no caso das plataformas Watson Assistant e Dialogflow (86,67%). Quanto à plataforma Alexa, o percentual médio de acerto na classe Específica foi de 75%. No entanto, a análise inferencial não apontou diferenças estatísticas entre os percentuais de acerto das três plataformas.

Sendo assim, a plataforma que obteve maior capacidade global de generalização em seu modelo cognitivo para esse experimento foi a plataforma do Watson Assistant da IBM, seguida da plataforma Alexa Developer Console da Amazon e, por último, o Dialogflow do Google.

Como limitações deste estudo, pode-se citar que somente foi avaliado o modelo cognitivo, considerando apenas alguns casos específicos e com as versões disponíveis das plataformas no momento da realização do teste (ano de 2020).

9 REFERÊNCIAS

10 WONDERFUL EXAMPLES OF USING ARTIFICIAL INTELLIGENCE (AI) FOR GOOD, 2020. **Forbes**, junho. 2020. Disponível em:

<<https://www.forbes.com/sites/bernardmarr/2020/06/22/10-wonderful-examples-of-using-artificial-intelligence-ai-for-good>>. Acesso em: 10 de abril de 2021.

ADOPTING THE POWER OF CONVERSATIONAL UX | CHATBOTS. **Deloitte UK**, 2017. Disponível em:< <https://www2.deloitte.com/content/dam/Deloitte/nl/Documents/financial-services/deloitte-nl-fsi-chatbots-adopting-the-power-of-conversational-ux.pdf> >. Acesso em: 10 de abril de 2021.

BARR, Avron. FEIGENBAUM, Edward A. **The Handbook of Artificial Intelligence**, 1 Ed. Department of Computer Science Stanford University, California, United States of America, 1981.

BERGER, Renner. EBNER, Markus. EBNER, Martin. Conception of a Conversational Interface to Provide a Guided Search of Study Related Data, **International Journal of Emerging Technologies in Learning**, Graz University of Technology, Graz, Austria, 2019. Disponível em: < <https://online-journals.org/index.php/i-jet/article/view/10137/5602>>. Acesso em: 10 de abril de 2021.

CANÇADO, Márcia. **Manual de semântica: noções básicas e exercícios**. 2. Ed. Belo Horizonte Editora UFMG, 2008.

CREVIER, Daniel. **Ai: The Tumultuous History Of The Search For Artificial Intelligence**, BasicBooks, New York, United States of America ,1993.

DIAGRAMA DO FLUXO BÁSICO PARA CORRESPONDÊNCIA DE INTENÇÕES E RESPOSTA AO USUÁRIO FINAL. Disponível em:

<<https://cloud.google.com/dialogflow/docs/intents-overview>>. Acesso em: 02 de abril de 2021.

DIALOGFLOW. In **WIKIPÉDIA: a enciclopédia livre**. Disponível em:

<<https://en.wikipedia.org/wiki/Dialogflow>>. Acesso em: 02 de abril de 2021.

EISENSTEIN, Jacob. **Natural Language Processing**, MIT Press, 2018. Disponível em:<<https://github.com/jacobeisenstein/gt-nlp-class/blob/master/notes/eisenstein-nlp-notes.pdf>>. Acesso em: 02 de abril de 2021.

ESCOLA DE ROBÔ BANCO DO BRASIL, 2019. Disponível em:

<https://www.bb.com.br/pbb/pagina-inicial/imprensa/n/60290/Banco%20do%20Brasil%20inaugura%20Escola%20de%20Rob%C3%B4#>. Acesso em: 02 de abril de 2021. ¹

GARTNER HYPE CYCLE DE INTELIGÊNCIA ARTIFICIAL, 2019. Disponível em:

<<https://www.gartner.com/smarterwithgartner/top-trends-on-the-gartner-hype-cycle-for-artificial-intelligence-2019/>>. Acesso em: 02 de abril de 2021.

GOOGLE TRENDS, sobre o termo “Inteligência Artificial” nos últimos 5 anos Brasil.

Disponível em:<<https://trends.google.com.br/trends/explore?date=today%205-y&geo=BR&q=Intelig%C3%A2ncia%20Artificial>>. Acesso em: 02 de abril de 2021.

INTENÇÕES IBM, 2021. Disponível em:

<<https://cloud.ibm.com/docs/assistant?topic=assistant-intents&locale=pt-BR> janeiro de 2021>.
Acesso em: 02 de abril de 2021.

KHAN, Rashid; DAS, Anik. **Build Better Chatbots: A Complete Guide to Getting Started with Chatbots**. Editora Apress, Bangalore, Karnataka, India, 2017.

KLEINBAUM, David; KLEIN, Mitchel. **Logistic Regression - A Self-Learning Tex**. Springer, Emory University, Atlanta, GA, USA, 2010.

LUGER, George F. **Inteligência Artificial**, 6 Ed. Pearson, São Paulo, Brasil, 2013.

MAPA DO ECOSISTEMA BRASILEIRO DE BOTS, 2020. Disponível em:

<<https://www.mobiletime.com.br/pesquisas/mapa-do-ecossistema-brasileiro-de-bots-2020/>>.
Acesso em: 02 de abril de 2021.

MCCORDUCK, Pamela. **Machines Who Think: A Personal Inquiry into the History and Prospects of Artificial Intelligence**, A K Peters, Massachusetts, USA, 2014.

MONTGOMERY, Douglas C.; RUNGER, George C. **Estatística Aplicada e Probabilidade para Engenheiro**, 5 Ed. LTC, 2012.

MRV LANÇA NOVA ASSISTENTE VIRTUAL, MIA, 2020. Disponível em:

<<https://www.mrv.com.br/institucional/pt/relacionamentos/noticias/mrv-lanca-nova-assistente-virtual-para-aprimorar-a-experiencia-dos-seus-clientes>>. Acesso em: 02 de abril de 2021. ²

NLP IBM, 2021. Disponível em:< <https://www.ibm.com/watson/natural-language-processing>>.
Acesso em: 02 de abril de 2021.

NLP, Sathiyakugan, 2018. Disponível em: <<https://devopedia.org/natural-language-processing>>. Acesso em: 02 de abril de 2021.

RUSSELL, Stuart. NORVIG, Peter. **Inteligência Artificial**. Trad: MACEDO, Regina. Elsevier, Rio de Janeiro, Brasil, 2013.

TROMP, John. FARNEBACK, Gunnar. **Combinatorics of Go**. Computers and Games: 5th International Conference, Springer, Turin, Italy, 2006.

TURING, Alan. **Computing Machinery and Intelligence**, Oxford University, UK, 1950.

Disponível em: <<https://academic.oup.com/mind/article/LIX/236/433/986238>>. Acesso em: 11 de abril de 2021.