

FACULDADES DOCTUM DE CARATINGA

CLEITON MESSIAS DA SILVA LOPES

**ANÁLISE DE CHATBOT APLICADO À QUESTÕES DE AVALIAÇÕES
ESCOLARES**

CARATINGA

2019

CLEITON MESSIAS DA SILVA LOPES

FACULDADES DOCTUM DE CARATINGA

**ANÁLISE DE CHATBOT APLICADO À QUESTÕES DE AVALIAÇÕES
ESCOLARES**

Trabalho de Conclusão de Curso apresentado ao curso de Ciência da Computação das Faculdades Doctum de Caratinga, como requisito parcial à obtenção do título de Bacharel em Ciência da Computação.
Área de concentração: Redes Neurais.
Orientador: Prof. Esp. Maicon Vinicius Ribeiro.

CARATINGA

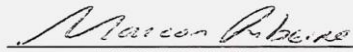
2019

TERMO DE APROVAÇÃO

O Trabalho de Conclusão de Curso intitulado: ANÁLISE DE CHATBOT APLICADO A QUESTÕES DE AVALIAÇÕES ESCOLARES, elaborado pelo(s) aluno(s) CLEITON MESSIAS DA SILVA LOPES foi aprovado por todos os membros da Banca Examinadora e aceito pelo curso de CIÊNCIA DA COMPUTAÇÃO das FACULDADES DOCTUM DE CARATINGA, como requisito parcial da obtenção do título de

BACHAREL EM CIÊNCIA DA COMPUTAÇÃO.

Caratinga 04/12/2019



MAICON RIBEIRO
Prof. Orientador



FABRÍCIA PIRES
Prof. Avaliador 1



RICARDO BOTELHO
Prof. Examinador 2

RESUMO

Os *chatbots* tiveram um impacto notável nas indústrias de tecnologia. Utilizando os avanços atuais da Inteligência Artificial, os *chatbots* são ferramentas que simulam conversas para fornecer assistência e aprendizado. A maioria das pessoas provavelmente já interagiu com um *chatbot* em algum momento, sendo as formas mais prováveis por meio de alguns assistentes de voz (como Siri, Amazon Alexa, Google Home ou Microsoft Cortana) ou através de aplicativos de mensagens (como o Facebook Messenger). Sendo assim, é possível encontrar várias áreas que já utilizam *chatbots* para automatizar seus processos. Partindo deste ponto de vista, tem-se a dúvida de como seria a eficiência de *chatbots* aplicados no meio acadêmico, com a finalidade de auxiliar alunos no momento de fazer a revisão de seu conhecimento por meio de questões direcionadas ao *chatbot*. Quais aplicativos possíveis poderiam atender à esta demanda nas escolas? Pode-se utilizá-los como ferramentas de aprendizado na sala de aula? Podem auxiliar a sanar as dúvidas dos alunos? Contudo, esse trabalho teve como intuito o desenvolvimento de um chatbot utilizando redes neurais no modelo seq2seq a fim de analisar se um *chatbot* é eficaz para atuar na área escolar. que foi treinado com 10 disciplinas diferentes, onde 80% das questões foram utilizadas para o treinamento da rede neural e 20% para realização dos testes, e com isso constatou-se que a média de assertividade do chatbot foi de 27%.

Palavras-chave: RNA. Inteligência Artificial. Chatbot. Disciplinas.

ABSTRACT

Chatbots have had a noticeable impact on the technology industries. Using today's advances in Artificial Intelligence, chatbots are conversational simulation tools to provide assistance and learning. Most people have probably interacted with a chatbot at some point, the most likely ways being through some voice assistants (such as Siri, Amazon Alexa, Google Home or Microsoft Cortana) or through messaging applications (such as Facebook Messenger).). Thus, it is possible to find several areas that already use chatbots to automate their processes. From this point of view, one wonders what the efficiency of chatbots applied to academia would be like, in order to assist students in reviewing their knowledge through questions directed to chatbot. What possible applications could meet this demand in schools? Can they be used as learning tools in the classroom? Can they help answer students' questions? However, this work had as intuition the development of a chatbot using neural networks in the seq2seq model in order to analyze if a chatbot 4907050 is effective to work in the school area. who was trained in 10 different subjects, where 80% of the questions were used for neural network training and 20% for testing, and it was found that chatbot's average assertiveness was 27%.

Key words: RNA. Artificial intelligence. Chatbot. Subjects.

LISTA DE SIGLAS

- CSV - Comma-separated values (Valores separados por delimitador)
- ENEM – Exame Nacional do Ensino Médio
- GHz - *Giga-hertz*
- HTML- *Hypertext Markup Language*
- IA - Inteligência Artificial
- LSTM - *Long Short Term Memory* (Memória de Longo Prazo de Curta Duração)
- PHP - *Hypertext Preprocessor*
- PLN - Processamento de Linguagem Natural
- RAM - Random Access Memory (Memória de acesso aleatório)
- RNA - Rede Neural Artificial
- RNAs - Redes Neurais Artificiais
- RNR - Rede Neural Recorrente
- Seq2seq - *Sequence to Sequence* (Sequência para Sequência)
- SQL - *Structured Query Language* (Linguagem de Consulta Estruturada)
- SSD - *Solid-state Drive* (Unidade de Estado Sólido)
- SSH - *Secure Shell* (Capsula segura)

LISTA DE FIGURAS

Figura 1: Processo de conversação entre o <i>chatbot</i> e o humano	16
Figura 2: Comparação entre RNR e rede neural <i>feedforward</i>	19
Figura 3: Processo do <i>chatbot</i>	22
Figura 4: Estrutura do projeto.....	23
Figura 5 : Tabelas	24
Figura 6: Questão e alternativas	26
Figura 7: Arquivo dataset_80.txt.....	27
Figura 8: Dataset_20.txt	28
Figura 9: Processo de <i>tokenização</i>	29
Figura 10: Estrutura <i>seq2seq</i>	30
Figura 11: Estrutura do projeto após a realização do treinamento	33
Figura 12: Estrutura do projeto após a realização do teste	34

LISTA DE GRÁFICOS

Gráfico 1: Épocas e assertividades na disciplina de português	32
Gráfico 2: Acertos e erros (disciplina de português).....	36
Gráfico 3: Similaridade resposta maior e similaridade resposta certa (disciplina de português)	37
Gráfico 4: Frequência de ocorrências das similaridades entre a resposta do <i>chatbot</i> e a alternativa que tem maior similaridade (disciplina de português)	38
Gráfico 5: Frequência de ocorrências das similaridades entre a resposta do <i>chatbot</i> e a alternativa correta (disciplina de português).....	39
Gráfico 6: Acertos e erros (disciplina de medicina)	40
Gráfico 7: Similaridade resposta maior e similaridade resposta certa (disciplina de medicina).....	41
Gráfico 8: Frequência de ocorrências das similaridades entre a resposta do <i>chatbot</i> e a alternativa que tem maior similaridade (disciplina de medicina).....	42
Gráfico 9: Frequência de ocorrências das similaridades entre a resposta do <i>chatbot</i> e a alternativa correta (disciplina de medicina)	43
Gráfico 10: Acertos e erros (disciplina de medicina)	44
Gráfico 11: Similaridade resposta maior e similaridade resposta certa (disciplina de medicina).....	45
Gráfico 12: Frequência de ocorrências das similaridades entre a resposta do <i>chatbot</i> e a alternativa que tem maior similaridade (disciplina de medicina).....	47
Gráfico 13: Frequência de ocorrências das similaridades entre a resposta do <i>chatbot</i> e a alternativa correta (disciplina de medicina)	48
Gráfico 14: Acertos e erros (disciplina de medicina)	49
Gráfico 15: Similaridade resposta maior e similaridade resposta certa (disciplina de medicina).....	50
Gráfico 16: Frequência de ocorrências das similaridades entre a resposta do <i>chatbot</i> e a alternativa que tem maior similaridade (disciplina de medicina).....	52
Gráfico 17: Frequência de ocorrências das similaridades entre a resposta do <i>chatbot</i> e a alternativa correta (disciplina de medicina)	53
Gráfico 18: Acertos e erros por disciplinas	54
Gráfico 19: similaridade por disciplinas	55
Gráfico 20: Acertos e erros com similaridades	56

LISTA DE QUADROS

Quadro 1: Estrutura do <i>dataset</i>	25
Quadro 2: Disciplinas	31
Quadro 3: result.csv	35

SUMÁRIO

1 INTRODUÇÃO	10
2 REFERENCIAL TEÓRICO	12
2.1 Aprendizagem	12
2.2 Avaliação	12
2.2.1 Avaliação escolar	13
2.3 Chatbot	14
2.3.1 Definição	14
2.4 Processamento de Linguagem Natural (PLN)	17
2.5 Redes Neurais Artificiais (RNAs)	18
2.6 Tipos redes neurais para processamento de linguagem natural	18
2.6.1 Recurrent Neural Network (RNR)	18
2.6.2 Long Short-Term Memory (LSTM)	19
2.6.3 Sequence to Sequence Models (seq2seq)	20
2.7 TensorFlow	20
2.8 Deep learning	20
2.9 Inteligência Artificial (IA)	21
2.10 SpaCy	21
3 METODOLOGIA	22
3.1 Estrutura do projeto	22
3.2 Dataset	24
3.2.1 Limpeza do <i>dataset</i>	26
3.2.2 Divisão <i>dataset</i>	27
3.2.3 Pré-processamento	28
3.3 Definição do modelo Seq2Seq	29
3.3.1 Desenvolvimento	30

3.4 Treinamento	31
3.5 Testes	33
4 ANÁLISE DOS RESULTADOS	36
4.1.1 Português	36
4.1.2 Medicina	40
4.1.1 Informática.....	44
4.1.2 Psicologia	49
4.1.3 Demais disciplinas.....	53
4.1.4 Análise geral das disciplinas	54
5 CONCLUSÃO	57
6 TRABALHO FUTURO	58
REFERÊNCIAS	59
APÊNDICE A	63
APÊNDICE B	82

1 INTRODUÇÃO

Atualmente *chatbots* são utilizados em várias plataformas, como por exemplo: no aplicativo de mensagens do *Facebook*, no *Slack*, *WhatsApp*, nos sites de empresas, entre outros. Em poucas palavras, um *chatbot* é basicamente um robô capaz de fornecer comunicação automatizada. Ou seja, é um programa treinado para conversar com pessoas como se fosse um ser humano. O objetivo principal da maioria dos *chatbots* é assumir algumas das atividades de trabalho de um profissional da área de atendimento ao cliente.

Em geral, os *chatbots* são classificados em dois tipos, os que respondem a perguntas padrões e os que utilizam IA (Inteligência Artificial), esses são melhores por conseguir respostas para perguntas nunca antes vistas, porém são mais custosos para serem desenvolvidos.

Com os surgimentos de vários *chatbots*, atuando em áreas diferentes, gerou-se uma dúvida, sendo esta, a ideia de poder utilizar um *chatbot* para responder perguntas de avaliação escolar.

Os *chatbots* podem ser uma ferramenta inovadora no processo de ensino aprendizagem, como por exemplo: ser um professor online 24 horas para tirar dúvidas de alunos. Quando ele é alimentado com as informações sobre o conteúdo disciplinar, os alunos podem aproveitar o momento de estudo para tirar dúvidas com o robô.

Visto isso, o presente trabalho teve como finalidade a criação de um *chatbot* com rede neural recorrente, alimentado com informações relacionadas a avaliação escolar. O objetivo é analisar a porcentagem de acertos e erros, caso o mesmo tenha uma boa porcentagem de acertos, pode-se utilizá-lo no meio acadêmico. Por exemplo: como uma forma de ajuda para alunos quando tiverem dúvidas sobre determinadas questões.

Essa pesquisa concentra em um *chatbot* com o propósito alternativo, um *chatbot* com Inteligência Artificial utilizando o modelo *seq2seq* (Sequência para Sequência) do *TensorFlow* com o propósito de responder a questões de avaliação escolar, de forma que, ao enviar as questões, o *chatbot* fez uso de sua rede neural juntamente com PLN (Processamento de Linguagem Natural) para assim responder as questões a ele submetidas.

Sendo assim, o presente trabalho propôs o desenvolvimento de um *chatbot* que foi treinado para essa finalidade. O principal objetivo da pesquisa é analisar se o comportamento de *chatbot* para responder avaliação escolar é eficaz ou não.

Em primeiro lugar, serão abordados conceitos utilizados para o desenvolvimento de um *chatbot*, como técnicas de aprendizagem e tipos de redes neurais. Depois disso será apresentada a estrutura do conjunto de dados. Após isso os dados serão treinados pela rede neural, e novos dados (questões nas quais a rede neural não foi treinada) serão submetidos ao *chatbot*. Por fim, pode-se passar para o processo de análise da ferramenta, apresentando por meio de gráficos a porcentagem de acertos e erros do *chatbot*, com o objetivo de analisar o uso de *chatbot* para responder questões fechadas (múltipla escolha) de avaliação escolar. De uma forma simples os passos seguidos foram os seguintes:

- Revisão bibliográfica de livros, artigos, monografias dentre outros relacionados à avaliação escolar, *chatbot*, inteligência artificial, rede neural e *TensorFlow*.
- Formatação dos dados do *dataset*, desenvolvimento do *chatbot*, treinamento com questões escolares.
- Submeter o *chatbot* a questões inexistentes no seu treinamento, analisando se o mesmo é eficaz para essa finalidade e apresentar os resultados.
- Analisar os resultados

Sendo assim, o *chatbot* foi testado com intuito de analisar se seu uso pode ser eficaz na área educacional, e para isso, no próximo tópico será explicado os principais conceitos utilizados no decorrer da pesquisa.

2 REFERENCIAL TEÓRICO

Este capítulo apresenta o resumo da pesquisa de livros, artigos, monografias, e outros, como meios de embasamento para o desenvolvimento deste trabalho.

2.1 Aprendizagem

Aprendizagem segundo Nunes (2017), “é o processo de aquisição e assimilação, mais ou menos consciente, de novos padrões e novas formas de perceber, ser, pensar e agir”.

Segundo Vygotsky (2004), “É como se aprendizagem colhesse os frutos do amadurecimento da criança, mas em si mesma a aprendizagem continua indiferente ao desenvolvimento.”

2.2 Avaliação

A avaliação é muito importante para o ensino-aprendizagem, pois através dela, é possível medir o nível de rendimento do aluno, para assim conseguir entender como está o desenvolvimento dos alunos.

Segundo Luckesi (2005):

O termo avaliar também tem sua origem no latim, provindo da composição a-valere, que quer dizer “dar valor a...”. Porém, o conceito “avaliação” é formulado a partir das determinações da conduta de “atribuir um valor ou qualidade a alguma coisa, ato ou curso de ação...”, que, por si, implica um posicionamento positivo ou negativo em relação ao objeto, ato ou curso de ação avaliado. Isto quer dizer que o ato de avaliar não se encerra na configuração do valor ou qualidade atribuídos ao objeto em questão, exigindo uma tomada de posição favorável ou desfavorável ao objeto de avaliação, com uma conseqüente decisão de ação. (p. 47).

Sendo assim, um elemento integrante do processo de ensino-aprendizagem, ganhou um espaço muito amplo nas escolas, como supracitado, avaliação é a forma de atribuir valor ou qualidade, dessa forma é possível saber o rendimento das pessoas em determinada área através do processo de avaliação, visto isso é possível aplicar avaliações nas escolas com intuito de medir o nível de aprendizado dos alunos em sala de aula.

2.2.1 Avaliação escolar

A avaliação desempenha um papel importante nas escolas. Ajuda professores e alunos a melhorar o ensino e a aprendizagem.

Como afirma Caldeira (2000):

A avaliação escolar é um meio e não um fim em si mesma; está delimitada por uma determinada teoria e por uma determinada prática pedagógica. Ela não ocorre num vazio conceitual, mas está dimensionada por um modelo teórico de sociedade, de homem, de educação e, conseqüentemente, de ensino e de aprendizagem, expresso na teoria e na prática pedagógica. (p. 122).

Pode-se mencionar às práticas de avaliação sob a forma de exames e provas, utilizadas em colégios católicos da Ordem Jesuítica e em escolas protestantes, a partir do século XVI. Conforme Luckesi (2003):

A tradição dos exames escolares, que conhecemos hoje, em nossas escolas, foi sistematizada nos séculos XVI e XVII, com as configurações da atividade pedagógica produzidas pelos padres jesuítas (séc. XVI) e pelo Bispo John Amós Comênio (fim do séc. XVI e primeira metade do século XVII). (p. 16).

De forma geral, a avaliação escolar pode ser definida como um meio de obter informações sobre os avanços e as dificuldades de cada aluno, constituindo-se em um procedimento permanente de suporte ao processo ensino-aprendizagem, de orientação para o professor planejar suas ações, a fim de conseguir ajudar o aluno a prosseguir, com êxito, seu processo de escolarização. Os instrumentos de avaliação mais usados são provas escritas ou orais, seminários, tarefas, pesquisas e dinâmicas de grupos (MENEZES, 2001).

Visto isto, a avaliação é uma forma de obter informações sobre o rendimento dos alunos no processo de ensino-aprendizagem, pois ela fornece um feedback útil para instrutores e alunos sobre o grau em que os alunos estão alcançando com sucesso os objetivos de aprendizagem do curso, permitindo assim que os instrutores determinem as métricas de medida para a compreensão e proficiência dos alunos nos objetivos de aprendizagem do curso.

Muitas vezes os professores passam uma revisão antes de aplicar a avaliação com o objetivo de relembrar os alunos das matérias, porém alguns alunos por medo, vergonha, ou outros motivos, acabam não conseguindo tirar dúvidas com os

professores. Segundo Newway (2017), ‘bot de revisão’ ou ‘tira-dúvidas’ serve de reforço para os conteúdos trabalhados, especialmente em períodos que antecedem provas, vestibulares ou o ENEM.

2.3 Chatbot

Neste capítulo será descrito a definição de um *chatbot*, já que partes essenciais desse trabalho se baseiam nesse assunto. Depois de definir o que é um *chatbot*, a tecnologia usada para alimentar os *chatbots* é explicado.

2.3.1 Definição

A interação de fala ou texto entre o humano e o computador está ganhando mais e mais popularidade hoje em dia. As pessoas querem se comunicar com computadores da mesma maneira como eles se comunicam com outros humanos seres. Uma das principais ferramentas utilizadas para analisar a fala e fornecer respostas semelhantes a humanos são o Processamento de Linguagem Natural (PLN). A fim de fornecer respostas adequadas com base em frases ou palavras-chave tiradas de perguntas, bem como para manter a comunicação contínua, sendo assim os *chatbots* fazer uso dessa ferramenta para manter um diálogo com uma pessoa.

O termo *chatbot* consiste em dois outros termos: chat e bot. O significado pode ser melhor entendido examinando os dois componentes separadamente.

Para Infopédia (2003) chat é uma ferramenta de comunicação que permite a troca, em tempo real, de mensagens escritas entre dois ou mais utilizadores de uma rede de computadores. E segundo Loureiro (2016), a palavra “*bot*” vem de “*robot*”, que, em inglês, significa “robô”. Ou seja, um *bot* nada mais é do que um robô, mas que existe apenas em formato digital.

De acordo com Rocha (2018), um *chatbot* é definido da seguinte forma:

Chatbot é um programa de computador que utiliza inteligência artificial cada vez mais aperfeiçoada para imitar conversas com usuários de várias plataformas e aplicativos, como acontece no Facebook e em sites de e-commerce.

Em outras palavras, ele funciona como uma espécie de assistente que se comunica e interage com as pessoas através de mensagens de texto automatizadas.

Ou seja, as pessoas introduzem uma mensagem, e o programa deve fornecer a resposta inteligente mais adequada na forma de texto ou fala. Com o intuito de dar a melhor resposta para o usuário esses robôs passam por um processo de treinamento, onde são expostos os dados relacionados à área em questão.

Para Laven (2006), um *chatbot* é um programa de "Inteligência Artificial" que tenta simular uma conversa digitada, com o objetivo de enganar, pelo menos temporariamente, um humano a pensar que estava conversando com outra pessoa.

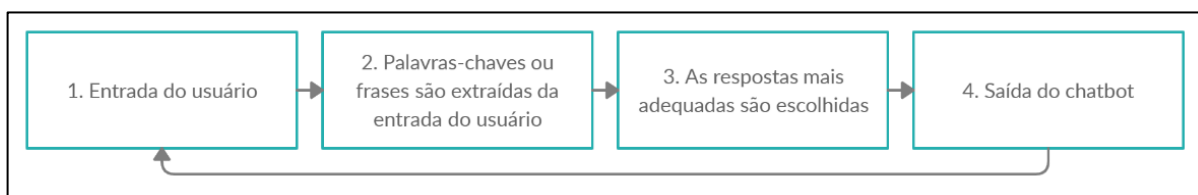
De acordo com Schappo (2018), *chatbot* é um programa de computador que faz o que é programado, simulando uma conversa humana em um chat. Dessa forma, é possível automatizar tarefas repetitivas e burocráticas, como dúvidas frequentes, na forma de diálogo pré-definido entre o usuário e um "robô".

O princípio subjacente a todo *chatbot* é interagir com um usuário humano (na maioria dos casos) por meio de mensagens de texto e se comportam como se fossem capazes de entender a conversa e responder adequadamente ao usuário. A origem dos computadores conversar com humanos é tão antiga quanto o próprio campo da ciência da computação.

Turing (1950) definiu um teste simples agora referido como o *teste de turing* em 1950 onde um juiz humano teria que prever se a entidade que eles estão se comunicando via texto é um programa de computador ou não. No entanto, o objetivo deste teste é muito maior que o caso de uso de *chatbots*, a principal diferença é que o conhecimento de domínio de um *chatbot* é limitado, enquanto o teste de Turing pressupõe que fale sobre qualquer tópico com o agente. Isso ajuda durante a criação de um *chatbot*, pois eles não precisam ter um conhecimento de domínio (potencialmente) infinito e, como tal, pode se concentrar em certos tópicos muito específicos, como por exemplo, ajudar os usuários reservam um quarto em um hotel.

Além disso, outra suposição geral que se tem ao desenvolver um *chatbot* é que os usuários normalmente têm uma meta que desejam alcançar até o final da conversa. Isso então influencia a conversa, com fluxos e tópicos para atingir o objetivo escolhido.

Portanto, a definição de um *chatbot* adotada para esta pesquisa é um programa de comunicação humanizada por texto e que presta serviços a usuários humanos para atingir uma meta bem definida, como pode ser visto na Figura 1.

Figura 1: Processo de conversação entre o *chatbot* e o humano

Fonte: O autor.

Segundo Lima (2018), existem basicamente 2 tipos de *chatbots*, sendo eles: *chatbots* baseados em comandos (regras), no qual não entendem perguntas que não sejam as previamente cadastradas. Ou seja, não podem criar novos textos. E *chatbots* baseados em IA (Inteligência Artificial), esses podem responder a perguntas ambíguas, isso significa que você não precisa ser específico ao fazer perguntas a esses *chatbots*, eles podem criar respostas a partir do zero usando PLN (Processamento de Linguagem Natural), podendo também aprender com seus erros e se tornar mais assertivo em suas respostas.

Para essa pesquisa foi escolhido o modelo baseado em IA, já que o mesmo faz uso de PLN conseguindo entender linguagem, contexto e intenção. Para entender melhor sobre os *chatbots* será descrito uma breve história dos mesmos.

A primeira instância de um agente de conversação nasceu em 1966: Eliza era um programa de computador que simulava um psiquiatra e reformulava a entrada do usuário usando (pelos padrões de hoje) técnicas de processamento de linguagem natural (WEIZENBAUM, 1966). Apesar de ser relativamente simples, o programa conseguiu dar a ilusão de entender os problemas do usuário e enganou com sucesso muitas pessoas.

Então, durante várias décadas, os *chatbots* seguiram fortemente a abordagem da Eliza com pequenas adições trazidas para o campo, como síntese de fala e gerenciamento de emoções. Então, em 2001, veio o *SmarterChild*, um agente de conversação desenvolvido pela *ActiveBuddy* (agora *Colloquis*) que operava no *AOL Instant Messenger* e *MSN Messenger*. Inspirado pelo surgimento de plataformas de mensagens instantâneas, como *SMS*, o *SmarterChild* foi criado para fornecer acesso rápido a notícias, previsões meteorológicas, esportes resultados, entre outros. A principal inovação foi que o *SmarterChild* estava conectado a uma base de conhecimento e detinha informações úteis para seus usuários. Infelizmente, existiam

muitas limitações técnicas do processamento de linguagem natural na época e eles foram “esquecidos” na história.

O próximo avanço para agentes de conversação foi feito por uma equipe da *IBM* através do projeto *Watson*. O agente foi projetado com o único objetivo de ganhar o programa de TV americano *Jeopardy*, que aconteceu em 2011 ao competir contra dois dos ex-campeões do programa. Infelizmente, esse agente só poderia responder a perguntas de uma linha e foi incapaz de manter uma conversa adequada com outra pessoa.

Finalmente, no início de 2010, surgiu a ascensão de assistentes virtuais como o *Siri* da *Apple*, *Cortana* da *Microsoft*, assistente do *Google*, *Alexa* da *Amazon* e outros. Essas agentes trouxeram em campo o conceito de conversação e diálogo orientado a objetivos. Outro grande evento no campo dos *chatbots* foi o lançamento da plataforma *Facebook Messenger* para o *Facebook* em 2016, e permitiu a criação de conversações agentes para empresas não relacionadas à IA.

Conforme mostrado neste breve resumo, houve muito progresso desde os primeiros dias dos agentes conversacionais. Isso não implica, no entanto, que as soluções atuais estão sem falhas.

2.4 Processamento de Linguagem Natural (PLN)

Segundo SAS (2019), o processamento de linguagem natural (PLN) é um ramo da inteligência artificial que ajuda os computadores a entender, interpretar e manipular a linguagem humana.

Boutin (2017), explica que a "intenção" de um usuário é capturada e classificada pelo bot. O bot pode então ser treinado com quantas tentativas forem necessárias, dependendo da finalidade do *chatbot*. Um exemplo de uma intenção poderia ser "conversa fiada". Um começaria a treinar por criando intenções como "como você está?", "é bom tempo hoje", ou "quando nos encontramos acima?". A PLN não usa palavras-chave, em vez disso, usa sua compreensão do reconhecimento de padrões, estruturas de frases e expressões idiomáticas para coincidir com a intenção dos usuários com o anteriormente classificado intenções

2.5 Redes Neurais Artificiais (RNAs)

Uma rede neural artificial é uma tentativa de simular a rede de neurônios que compõem o cérebro humano, para que o computador possa aprender coisas e tomar decisões de maneira humana. As RNAs são criadas programando computadores comuns para se comportarem como se fossem células cerebrais interconectadas. (MARR, 2018).

Segundo Dacombe (2017), rede neural artificial é um modelo matemático simples do cérebro usado para processar relações não lineares entre entradas e saídas em paralelo, como o cérebro humano faz a cada segundo.

Como os autores acima explicam as Redes Neurais Artificiais são uma representação do cérebro humano, no qual tem como finalidade se “comportar” como tal.

2.6 Tipos redes neurais para processamento de linguagem natural

A seguir será apresentado tipos de rede neurais para PLN (Processamento de Linguagem Natural).

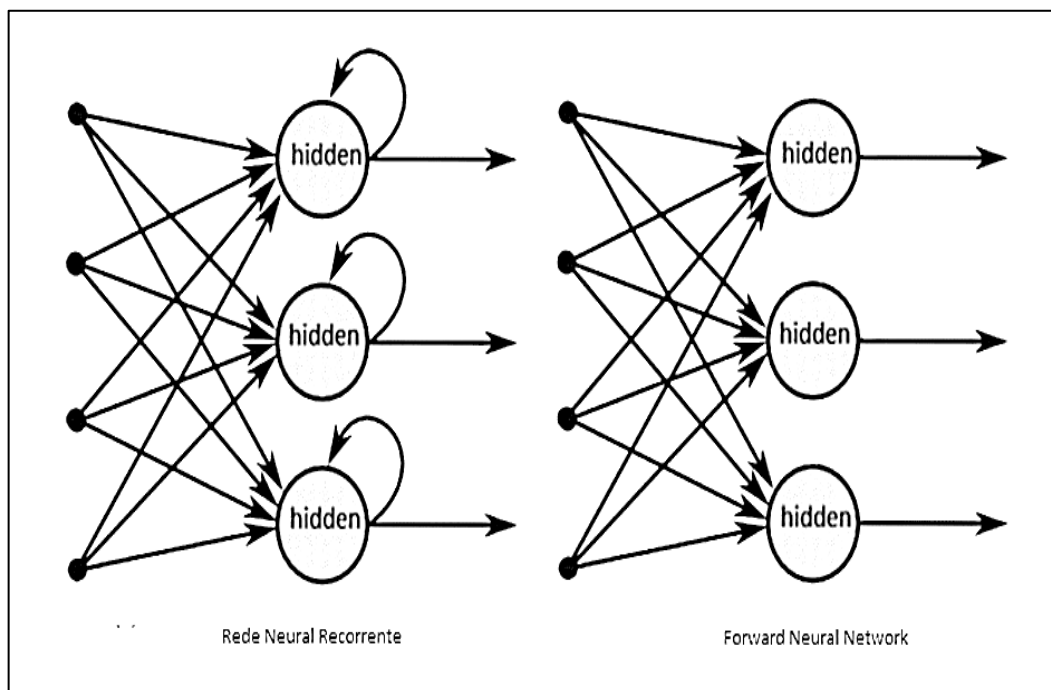
2.6.1 Recurrent Neural Network (RNR)

Segundo Davydova (2017), uma rede neural recorrente (RNR), diferentemente de uma rede neural *feedforward*, é uma variante de uma rede neural artificial recursiva, na qual, conexões entre neurônios fazem um ciclo dirigido. Isso significa que a saída depende não apenas das entradas atuais, mas também do estado do neurônio da etapa anterior. Essa memória permite que os usuários resolvam problemas de PNL.

Sendo assim, RNRs ou Redes Neurais Recorrentes são um tipo especial de redes neurais especificamente adaptado para sequências de dados. Em poucas palavras, são redes neurais com *loops*. Mais formalmente: É projetada para preservar o estado neuronal anterior, este permite que a rede neural retenha o contexto e produza resultados com base no estado anterior Davydova (2017). Essa abordagem torna as RNRs desejáveis para os *chatbots*, pois a manutenção do contexto em uma conversa é essencial para entender o usuário. RNRs são amplamente utilizados para tarefas de PNL, como tradução, fala reconhecimento, geração de texto e legenda de

imagens. A Figura 2 mostra a comparação entre a arquitetura RNRs e a rede neural *feedforward* (rede neural artificial em que as conexões entre os nós não formam um ciclo).

Figura 2: Comparação entre RNR e rede neural *feedforward*



Fonte: JEANS, 2013.

Como pode ser visto na Figura 2 a principal diferença entre a RNR (Rede Neural Recorrente) e a rede neural *feedforward*, é que, em cada neurônio da RNR, a saída do intervalo de tempo anterior é alimentada como entrada do próximo passo.

2.6.2 Long Short-Term Memory (LSTM)

De acordo com Davydova (2017), Long Short-Term Memory (LSTM) ou do português unidades de memória de longo e curto prazo é uma arquitetura de rede neural recorrente específica (RNN) projetada para modelar sequências temporais e suas dependências de longo alcance com mais precisão do que as RNNs convencionais.

As unidades de memória de longo e curto prazo (LSTMs) são amplamente utilizadas no Processamento de linguagem natural. Eles foram introduzidos pela primeira vez em 1997 por Hochreiter e Schmidhuber (2006) e se tornaram cada vez mais populares nos últimos anos, graças aos avanços no aprendizado acelerado por

hardware. Eles também mostraram resultados promissores em tradução automática e legendas de imagens.

2.6.3 Sequence to Sequence Models (seq2seq)

Para Davydova (2017), um modelo de sequência a sequência consiste em duas redes neurais recorrentes: um codificador que processa a entrada e um decodificador que produz a saída. O codificador e o decodificador podem usar os mesmos ou diferentes conjuntos de parâmetros. Esses modelos são usados principalmente em sistemas de resposta a perguntas, *chatbots* e tradução automática.

Como o nome sugere, *seq2seq* toma como entrada uma sequência de palavras (sentença ou sentenças) e gera uma sequência de saída de palavras. Isso é feito pelo uso da rede neural recorrente (RNN).

2.7 TensorFlow

Segundo Cavalheiro (2018) *TensorFlow* é uma biblioteca de software de código aberto para computação numérica usando gráficos de fluxo de dados, na qual foi originalmente desenvolvido pelo Google.

Conforme Yegulalp (2019) o *TensorFlow* foi criado pela equipe Google Brain (time de pesquisa de inteligência artificial no Google) e é uma biblioteca de código aberto compatível com Python para computação numérica que torna o aprendizado de máquina mais rápido e fácil.

Sendo assim o propósito desse sistema é ajudar pesquisadores e usuários a criar modelos para resolver PNL e reconhecimento de imagem usando bibliotecas de aprendizado de máquina e gráficos de fluxo de dados. Também pode ser usado no aprendizado profundo para criar arquitetura de rede neural e desenvolva algoritmos no aprendizado de máquina.

2.8 Deep learning

Uma vez que os *chatbots* imitam uma pessoa real (IA), técnicas são usadas para construí-las. Uma dessas técnicas dentro da IA é o *Deep Learning*, que imita o cérebro humano. Ele encontra padrões dos dados de treinamento e usa os mesmos padrões para processar novos dados.

2.9 Inteligência Artificial (IA)

A palavra inteligência vem do latim *inter* (entre) e *legere* (escolher). Na prática, inteligência é a habilidade de escolher, dentre várias possibilidades, aquela que permitirá realizar de forma eficiente uma determinada tarefa. A palavra artificial também vem do latim *artificiale* e represente algo não natural, isto é, algo feito pelo homem. Portanto, Inteligência Artificial (IA) é um tipo de inteligência, produzida pelo homem, para dotar as máquinas de algum tipo de habilidade que simule a inteligência do próprio homem.

Segundo Teixeira (1991), “A mente humana funciona como um computador, e por isso o estudo de programas computacionais é a chave para se compreender alguma coisa acerca de nossas atividades mentais.”.

Abaixo está listado duas definições de pesquisadores em inteligência artificial:

Segundo Margaret (2018), é a simulação de processos de inteligência humana por máquinas, especialmente sistemas de computador.

Segundo Copeland (2018), a capacidade de um computador ou controlado por computador robô para executar tarefas comumente associadas a seres inteligentes.

De modo geral a inteligência artificial é uma área da ciência da computação que enfatiza a criação de máquinas inteligentes que funcionam e reagem como seres humanos. Possibilita que as máquinas aprendam com a experiência, ajustem-se a novas entradas e realizam tarefas semelhantes às humanas.

2.10 SpaCy

Segundo Li (2018) *SpaCy* é uma biblioteca de processamento de linguagem natural popular e fácil de usar em Python. Ele fornece níveis atuais de precisão e velocidade de ponta e possui uma comunidade ativa de código aberto. Dentre as funções do *SpaCy*, podemos destacar a análise de similaridade, método este, que permite realizar uma comparação entre objetos, e determinar se existe uma possível semelhança entre eles, dado como exemplo, o cão que é muito parecido com um gato, enquanto que uma banana não é semelhante com nenhum deles. Os tokens idênticos são obviamente 100% semelhantes entre si.

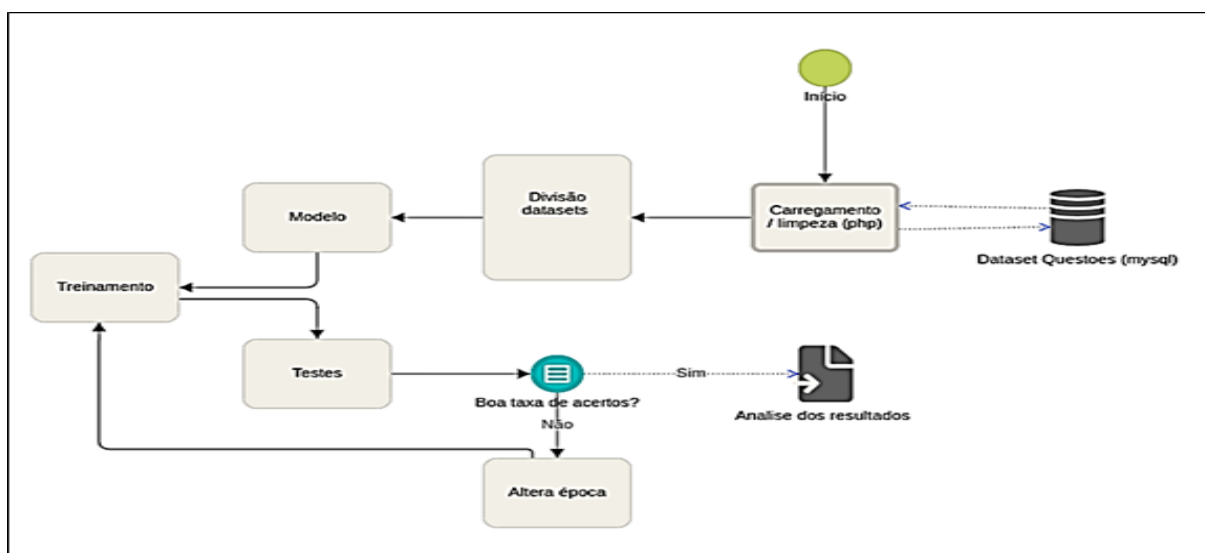
3 METODOLOGIA

A abordagem metodológica da pesquisa concentrou-se em criar um *chatbot* com *TensorFlow* utilizando redes neurais recorrentes, desenvolvido na linguagem de programação python, que foi treinado com questões escolares, e após, foi analisado a assertividade do *chatbot*, para assim, poder concluir se o mesmo pode ou não auxiliar os alunos.

3.1 Estrutura do projeto

A presente pesquisa foi separada em 5 etapas, sendo elas: preparação do dataset (onde ocorreu o processo de limpeza e a separação dos dados, sendo 80% para o treinamento da rede neural e 20 % para a realização dos testes), definição do modelo (o modelo utilizado foi o seq2seq, que consiste basicamente em duas redes neurais, a primeira é responsável por entender a pergunta e a segunda tem como finalidade retornar a resposta mais adequada para a pergunta), treinamento do *chatbot* (nessa etapa foram enviados os dados de treinamento contendo 80% das questões para a rede neural “aprender”), realização dos testes (foram enviados 20% das questões para o chatbot responde-las) e análise dos resultados. Na Figura 3 são mostrados os procedimentos nos quais ocorreram para o desenvolvimento dessa pesquisa.

Figura 3: Processo do *chatbot*

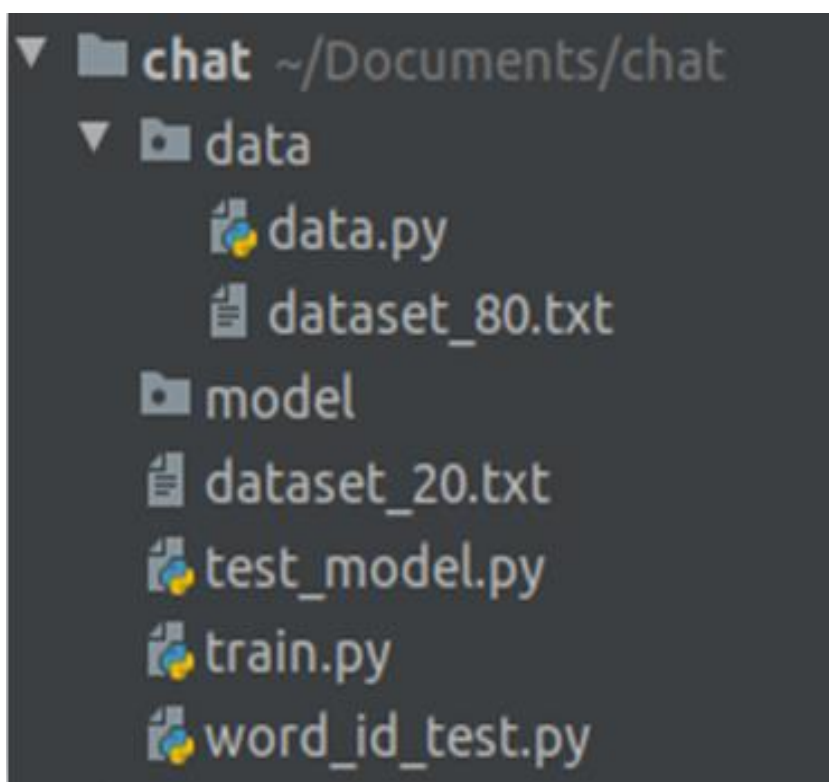


Fonte: O autor.

Pode ser visto na Figura 3 que inicialmente o *dataset* foi enviado para o processo de limpeza dos dados, em seguida foi feita a divisão e criação dos modelos, após isso foram realizados os treinamentos e testes. Para uma melhor performance foi realizado treinamentos com épocas diferentes (época basicamente é a quantidade de vezes que os dados são enviados para a rede neural aprender, como por exemplo a quantidade de vezes que os seres humanos estudam sobre determinado assunto com intuito aprender), a fim de identificar qual a melhor quantidade de época para o treinamento, e para finalizar foi realizado a análise dos resultados.

Para efetuar o desenvolvimento do *chatbot*, foi utilizada uma máquina com processador Intel Celeron de 1.83 GHz (*Giga-hertz*) com 4 núcleos, com 4 *gigabytes* de memória RAM (*Random Access Memory*), 480 *gigabytes* de SSD (*Solid-state Drive*), no sistema operacional Ubuntu na versão 18.04, utilizando a biblioteca *TensorFlow* na versão 1.5 para a linguagem Python na versão 7.2, a estrutura do projeto pode ser vista na Figura 4 e é explicada ao decorrer da metodologia.

Figura 4: Estrutura do projeto



Fonte: O autor.

Os arquivos com extensão *.py* que podem ser vistos na Figura 4 são *scripts* escritos na linguagem *python*.

3.2 Dataset

Um *chatbot* precisa de um *dataset*, que em uma definição simples, é um conjunto de dados, o qual deve ser - em relação ao presente estudo - submetido a treinamento por uma rede neural. A necessidade do *dataset* ocorre por dois motivos principais: saber o que as pessoas estão dizendo e saber o que responder, por exemplo: considere um simples *bot* de atendimento ao cliente. O *chatbot* precisa de uma ideia aproximada do tipo de perguntas que as pessoas vão fazer e, em seguida, precisa saber quais devem ser as respostas para essas perguntas. Para isso ele é treinado com *datasets* relacionado ao seu propósito.

O *dataset* (conjunto de dados) que foi utilizado para o treinamento do *chatbot*, inicialmente encontrava-se em um banco de dados *MySQL*, que foi disponibilizado pela empresa VersaTec que é uma empresa de Tecnologia, situada na cidade de Caratinga MG, fundada no dia 22 de novembro de 2012, em que a autorização do mesmo pode ser vista no APÊNDICE B. Todavia para o treinamento do *chatbot* os dados devem ter um fluxo de conversação. Deve haver uma sentença ou uma pergunta e uma resposta para assim, o *chatbot* conseguir entender o usuário e conseguir retornar uma resposta. Visto isto, foi realizado uma consulta na linguagem SQL (*Structured Query Language*), com a junção das tabelas representadas na Figura 5.

Figura 5 : Tabelas

questao	alternativas_questa	disciplina
id int(11)	id int(11)	id int(11)
banca_id int(11)	resposta text	descricao varchar(128)
usuario_id int(11)	label varchar(20)	
instituicao_id int(11)	correta varchar(1)	
questao text	questao_id int(11)	
pontos int(11)		
observacao text		
disciplina_id int(11)		
nivel_escolaridade_id int(11)		
materia_id int(11)		
texto_associado text		
reference int(11)		
conferida varchar(1)		
exibir varchar(1)		

Fonte: O autor.

Após a consulta SQL foram retornados os campos:

- Tabela 'questão'
 - Coluna 'questão' (contém as perguntas que foram utilizadas para o treinamento de teste)

- Tabela 'alternativas questão'
 - Coluna 'Reposta' (são as respostas das perguntas, sendo cinco alternativas para cada)
 - Coluna 'Correta' (indica se a resposta está certa ou errada, e somente uma das cinco respostas pode estar certa).

- Tabela 'disciplina'.9
 - Coluna 'descrição' (refere-se à qual disciplina determinada questão pertence).

Em seguida os dados foram exportados para um arquivo com extensão csv (valores separados por vírgula), criando assim o *dataset*, no qual tem a seguinte estrutura: questão, resposta, correta e disciplina conforme pode ser visto no Quadro 1.

Quadro 1: Estrutura do *dataset*

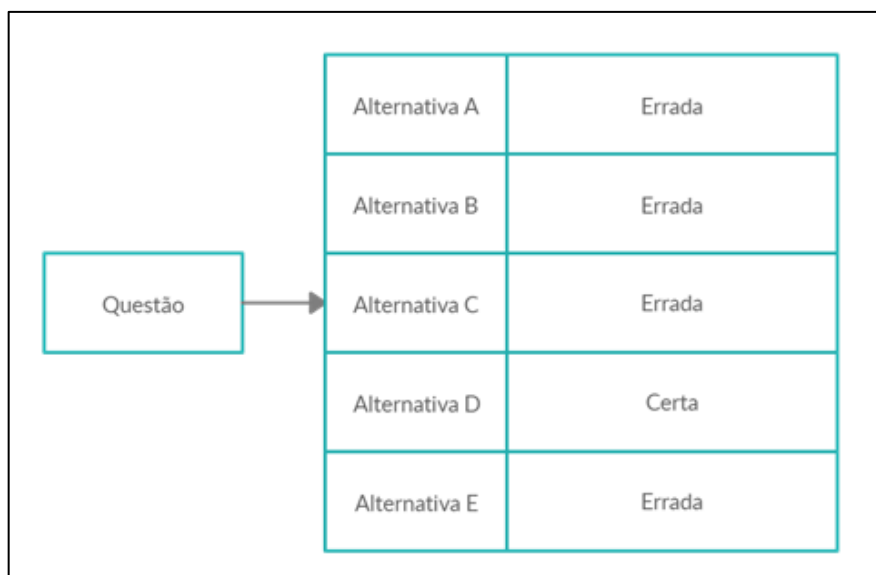
Colunas	Definição
questão	Descrição da questão
reposta	Alternativas da questão
correta	Sinaliza a resposta certa da questão
disciplina	Disciplina relacionada a questão

Fonte: O autor.

Para cada pergunta existem 5 alternativas, que são representadas pelas letras: A, B, C, D e E, onde somente uma está correta, conforme pode ser visto uma representação na figura Figura 6. Existem um total de 20.210 (vinte mil duzentos e dez) questões onde será aplicado 80% para o treinamento da rede neural e 20% para a realização dos testes, porém esses serão separados por disciplinas, para que o

chatbot possa ser treinado e testado em matérias diferentes, podendo assim ser feitas análises separadas.

Figura 6: Questão e alternativas



Fonte: O autor.

As informações incluídas nos *datasets* contêm tags na linguagem HTML (Linguagem de Marcação de Hipertexto) conforme pode ser visto no exemplo a seguir (as tags `<p>` e `</p>` pertencem à linguagem HTML, e significam aberturas e fechamentos, de parágrafos): `<p><p>`Na redação oficial, há três tipos de expedientes que se diferenciam mais pela finalidade que `</p><p>`pela forma: o ofício, o aviso e o memorando. Quanto a estes três expedientes, assinale a `</p><p>`alternativa CORRETA. `</p></p>`. Por esse motivo foi realizado um processo de limpeza no *dataset*, removendo tags HTML a fim de deixar os dados padronizados.

3.2.1 Limpeza do *dataset*

O processo de limpeza do *dataset* foi realizado por um *script* PHP (*Hypertext Preprocessor*), já que o mesmo fornece várias funções para manipulação de *strings* (sequência de caracteres), como por exemplo, a utilização de expressões regulares. A principal função utilizada foi `str_replace()`, que serviu para realização das substituições, na qual foi enviado 3 argumentos sendo eles: o valor a ser encontrado, o valor para substituir o valor na localização e pôr fim a string a ser pesquisada. Após o término da limpeza do *dataset*, passou-se então para a divisão do mesmo.

3.2.2 Divisão *dataset*

A separação do *dataset* foi realizada em duas fases, a primeira delas consistiu basicamente em dividir o *dataset* em 10 outros menores, sendo esses separados por disciplinas, afim de identificar o comportamento do chatbot em disciplinas diferente. Já a segunda fase consiste em dividir cada um dos novos *datasets* em dois conjuntos de dados distintos. O primeiro, contendo 80% das perguntas e respostas, para esse arquivo daremos o nome de `dataset_80.txt`, esses serão utilizados para o treinamento da rede neural, na Figura 7 pode ser visto um exemplo do seu conteúdo.

Figura 7: Arquivo `dataset_80.txt`

```
1 no aplicativo word da microsoft, por padrao, os comandos alinhamento, recuo e espacamento fazem parte da caixa de dialogo
2 paragrafo.
3 no aplicativo writer do libreoffice/broffice, por padrao, o comando imprimir pode ser acessado atraves do menu
4 arquivo.
5 mozilla firefox, internet explorer e google chrome sao exemplos de
6 navegadores de internet.
7 nos editores de texto word da microsoft e writer do libreoffice/broffice a funcao justificar
8 alinha o texto as margens esquerda e direita.
9 dentre os tipos de softwares existentes, podemos dizer que um programa de planilhas eletronicas e um:
10 software utilitario.
```

Fonte: O autor.

Os números representados no lado esquerdo da Figura 7 que vão de 1 a 10 são somente para uma visualização mais fácil, pois os mesmos não estão presentes no arquivo de *dataset*.

Pode ser visto que o `dataset_80.txt` segue a seguinte estrutura: pergunta, resposta, pergunta, resposta e assim sucessivamente, ou seja, para cada pergunta a sua resposta certa estará na linha seguinte, dessa forma todas as linhas ímpares são perguntas, e as pares são respostas.

O segundo conjunto de dados será utilizado para os testes, e esse contém 20% das perguntas e respostas, para esse daremos o nome `dataset_20.txt`, na Figura 8 pode ser visto seu conteúdo.

Figura 8: Dataset_20.txt

```
1 o termo anexo e utilizado pelos programas de correio eletronicos para designar
2 a prioridade da mensagem.
3 N
4 o arquivo enviado junto com a mensagem.
5 S
6 o destinatario da mensagem.
7 N
8 o assunto da mensagem.
9 N
10 o remetente da mensagem. conhecimentos especificos
11 N
12
13 no powerpoint 2010 br, o acionamento da tecla de funcao f5 tem por objetivo
14 formatar um grafico visando a comparar dados.
15 N
16 inserir uma tabela em um slide da apresentacao.
17 N
18 realizar a apresentacao a partir do primeiro slide.
19 S
20 configurar as transicoes para exibicao entre os slides.
21 N
22 escolher um design predefinido para a apresentacao de slides.
23 N
24
```

Fonte: O autor.

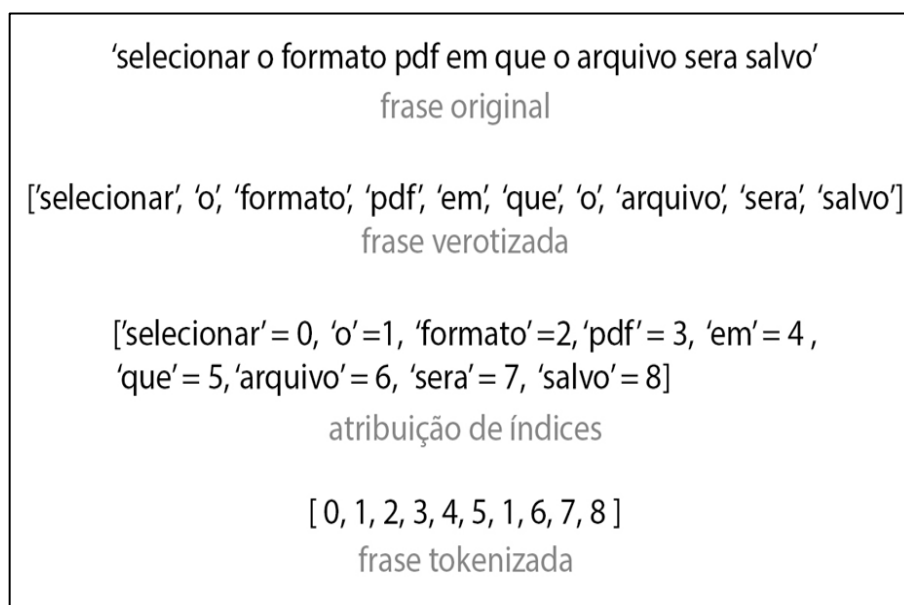
Vale a pena ressaltar que as perguntas do dataset_80.txt não estão presentes no dataset_20.txt, já que o mesmo é para realização dos testes. Para explicar melhor a estrutura do dataset_20.txt, será utilizado a representação das linhas (números no lado esquerdo) do arquivo, que pode ser visto na Figura 8: Na linha 1 encontra-se a questão, nas linhas 2, 4, 6, 8 e 10 tem-se as alternativas da questão (lembrando que cada questão tem um total de 5 alternativas: 'A', 'B', 'C', 'D' e 'E'), e nas linhas 3, 5, 7, 9 e 11 pode conter o valor S ou N, que são responsáveis por informar se a alternativa acima está certa (S) ou errada (N), sendo assim, somente pode conter um valor S para cada questão, e pôr fim a linha 12 representa o término da questão, esse processo ocorreu para todas as questões existentes no dataset_20.txt. Após a separação do *dataset* foi feito o pré-processamento dos dados, para que os mesmos pudessem ser enviados para a rede neural.

3.2.3 Pré-processamento

O arquivo data.py como pode ser visto na Figura 4, foi responsável por fazer o pré-processamento dos dados, de forma resumida: Foi lido o arquivo dataset_80.txt, todos os caracteres foram convertidos para *lower case* (minúsculas), após as frases

foram transformadas em vetores onde cada palavra é uma posição do vetor, após isso foi feito o processo de tokenização dos dados, que em uma definição simples, é a atribuição de Ids (Identificações Únicas) para as palavras como pode ser visto na Figura 9.

Figura 9: Processo de *tokenização*



Fonte: O autor.

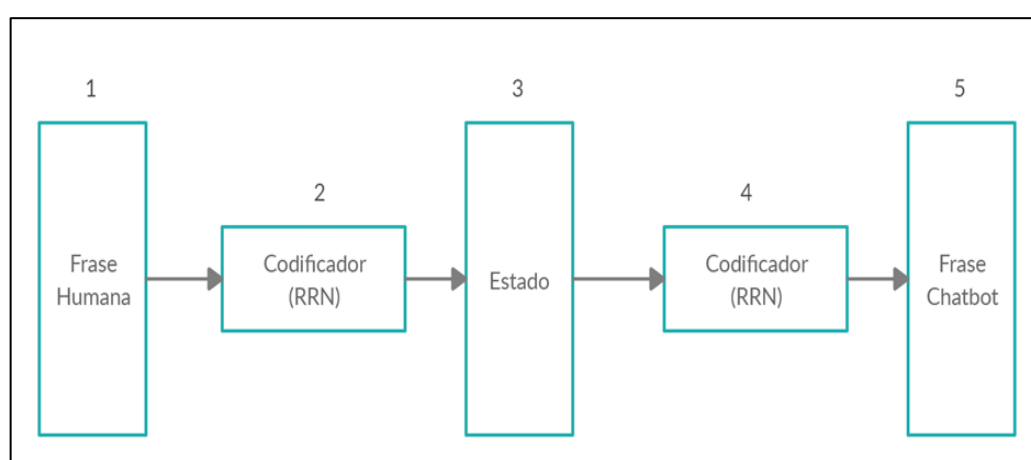
As frases são separadas em palavras, onde cada uma recebe um id (identificador único), sendo que palavras iguais terão o mesmo id, como pode ser visto na Figura 9 onde a letra 'o' recebe o id 1, e foi reiterado 2 vezes na frase.

3.3 Definição do modelo Seq2Seq

Com os dados devidamente formatados, passa-se então para a criação do modelo *Seq2Seq* que basicamente é como o chatbot irá funcionar, a estrutura do modelo pode ser vista na Figura 10. O codificador utiliza camadas de rede neural e converte as palavras de entrada em vetores ocultos correspondentes, cada vetor representa a palavra atual e o contexto da palavra. O decodificador é semelhante ao codificador. Toma como entrada o vetor oculto gerado pelo codificador, seus próprios estados ocultos e a palavra atual para produzir o próximo vetor oculto e finalmente prever a próxima palavra. Em outras palavras, codificador recebe os dados de entrada e os treina, depois passa o último estado de sua camada recorrente como um estado

inicial para a primeira camada recorrente da parte do decodificador. O decodificador acessa o último estado da última camada recorrente do codificador e o usa como um estado inicial para sua primeira camada recorrente, a entrada do decodificador é a sequência que espera-se obter, sendo assim, o codificador recebe como entrada a frase escrita por um humano e retorna um estado escondido para cada palavra, na qual é enviado para o decodificador, que é responsável por gerar a resposta do *chatbot*.

Figura 10: Estrutura *seq2seq*



Fonte: O autor.

Pode-se observar na Figura 10 que inicialmente o chatbot recebe uma frase como pergunta, a primeira rede neural é responsável por entender essa pergunta e gerar um estado (pergunta interpretada pela rede neural), e é enviado para a segunda rede neural, que é responsável por gerar a resposta e por fim tem-se a resposta da pergunta.

3.3.1 Desenvolvimento

O desenvolvimento foi realizado utilizando a linguagem python, pois nela existem muitas ferramentas que auxiliam em redes neurais, e o modelo *seq2seq* do *TensorFlow*, também conhecido como *Sequence-To-Sequence* (sequência para sequência). Esse modelo pode ser treinado para mapear uma sequência de entrada (perguntas) para uma sequência de saída (respostas), que não são necessariamente do mesmo tamanho que a outra. Com isso, tem-se o arquivo `train.py` que pode ser visto na Figura 4, que será utilizado para realizar o treinamento da rede neural.

3.4 Treinamento

Tendo o *script* de treinamento seguindo o modelo *seq2seq* e os *datasets* devidamente formatados, passou-se então para o procedimento de treinamento, no qual foi enviado para a rede neural os *dataset_80.txt*'s, nos quais contém 80% das perguntas, juntamente com as alternativas corretas, de modo que a rede neural possa aprender a responder questões semelhantes.

Foi executado o *script* *train.py* que pode ser visto na Figura 4. Como supracitado anteriormente, o treinamento foi realizado por disciplinas, nas quais foram escolhidas as que possuem maior quantidade de registros no banco de dados, pode ser observado no Quadro 2 as disciplinas e as quantidades totais de questões, as quantidades utilizadas para treinamento (80%), e pôr fim, as quantidades de questões utilizadas para a realização dos testes (20%).

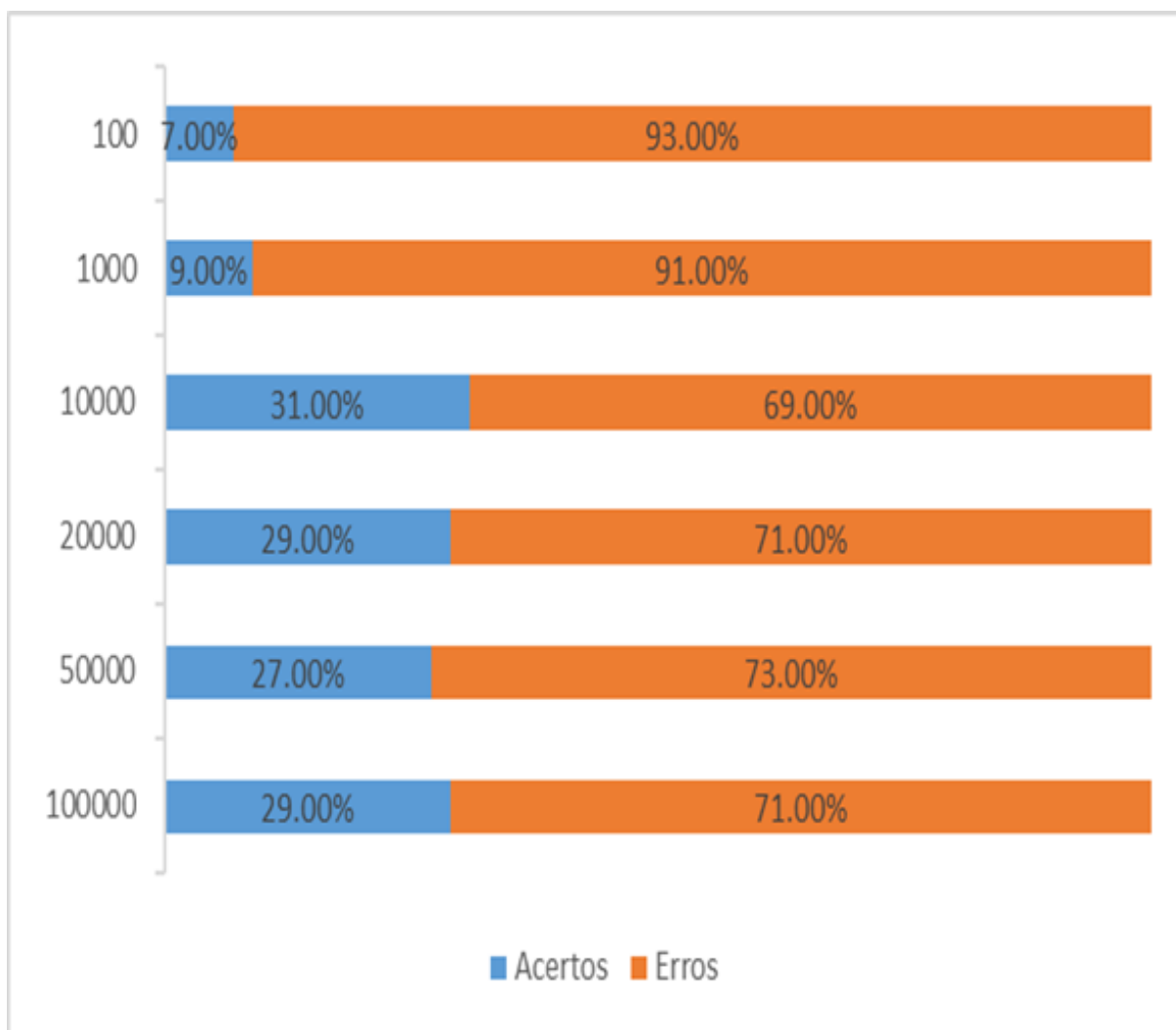
Quadro 2: Disciplinas

Matéria	Total Questões	Treinamento	Teste
Português	2440	1952	488
Informática	1745	1396	349
Medicina	2175	1740	435
Matemática	2655	2124	531
Pedagogia	1115	892	223
Psicologia	2125	1700	425
Enfermagem	1555	1244	311
Inglês	3400	2720	680
Raciocínio Lógico	1780	1424	356
Economia	1220	976	244

Fonte: O autor.

Cada *dataset* foi treinado por um total de 10.000 (dez mil) épocas, onde cada época corresponde a um ciclo de treinamento. Para chegar a esse resultado foram realizados teste com intuito de obter uma maior taxa de acertos, esses testes foram realizados utilizando a matéria de português e como pode ser visto no Gráfico 1, a melhor assertividade do *chatbot* foi quando treinado com 10.000 épocas.

Gráfico 1: Épocas e assertividades na disciplina de português



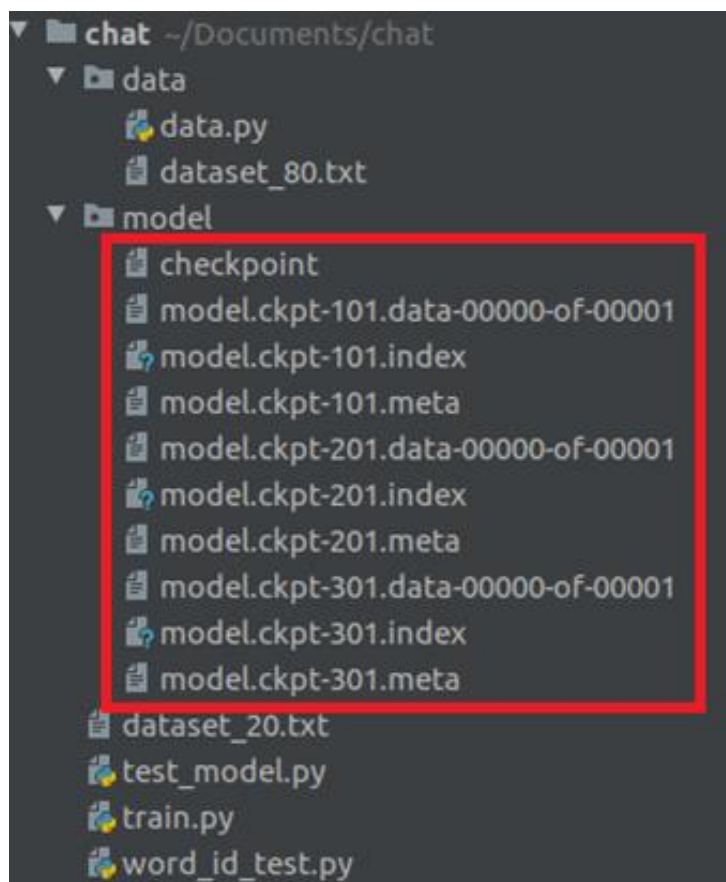
Fonte: O autor.

Ainda no Gráfico 1 pode-se observar que grandes números de épocas não significa que melhora a taxa de acertos, isso acontece pois a rede neural pode ficar viciada, tendo assim menor assertividade nas respostas.

Como citado no início da metodologia, o processador da máquina que estava à desenvolver a presente pesquisa era: processador *quad core* de 2 núcleos, e 4 *gb* de memória *ram*, sendo assim ela não conseguiu executar o treinamento, e para solucionar o problema foi criada uma instância no *google cloud* com as configurações: processador de 8 núcleos e 30 *gb* de memória, com isso foi possível realizar o treinamento com através de comandos *SSH (Secure Shell)*.

Após a finalização do treinamento, foram criados os arquivos de *checkingponit*, como pode-se observar na Figura 11 que é o projeto após a finalização do treinamento modelo já treinado.

Figura 11: Estrutura do projeto após a realização do treinamento



Fonte: O autor.

Os arquivos *checkpoint* serão utilizados para a realização dos testes, pois esses são os modelos já treinados. E finalmente passamos para a realização dos testes.

3.5 Testes

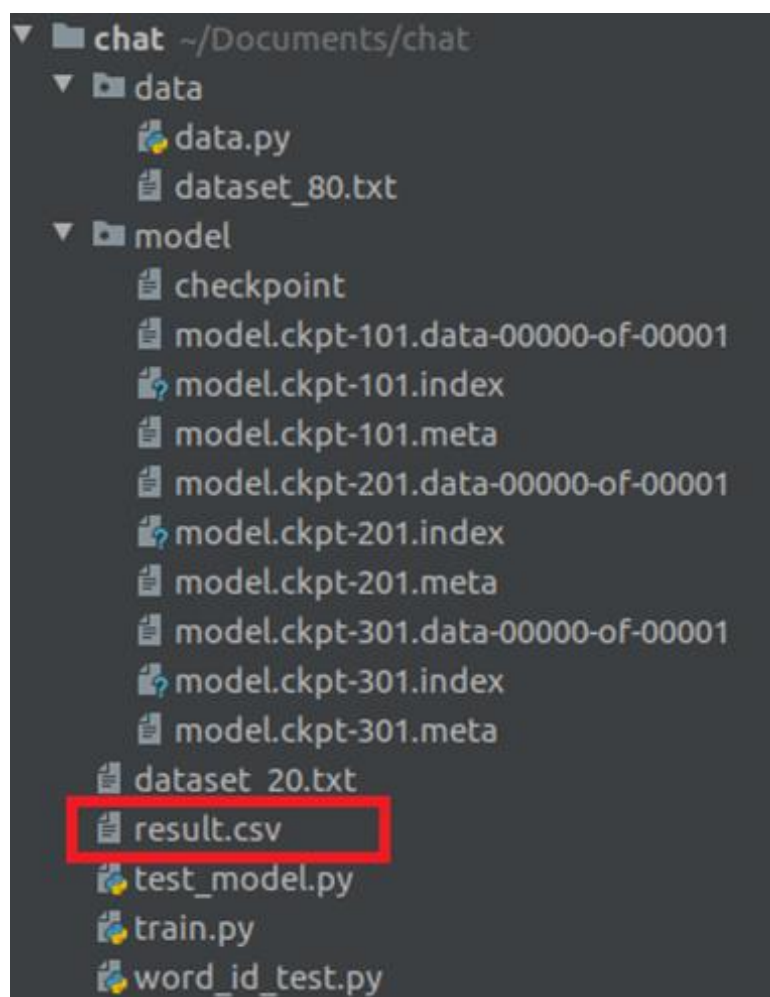
Após o treinamento, foi iniciado o processo de testes onde foram enviados os *datasets* *dataset_20.txt*'s nos quais contem 20%, das questões para o *chatbot* responder, e para isso foi utilizado o *script* com o nome *teste.py* que pode ser visto Figura 4.

A cada teste foi analisado a similaridade de resposta do *chatbot* com as 5 alternativas da questão. Para isso será utilizada a biblioteca *spacy* do *python*, o método *similarity* avalia a similaridade semântica estimada entre as palavras. Quando maior o valor, mais similar são as palavras, exemplo: cachorro e gato têm similaridade

de 0.877 enquanto cachorro e cassa tem similaridade de 0.673, ou seja, nesse exemplo cachorro e gato tem uma maior similaridade, sendo assim mais parecidos.

A variação da similaridade é entre -1 e 1, sendo que quanto mais próximo de 1, mais similar é o valor, e quanto mais próximo de -1 mais distante. Similaridades abaixo de 0 foram considerados respostas erradas, por ser muito diferente das alternativas, então a variação feita no teste foi entre 0 e 1, fim de deixar menos propício a 'chute' e melhorar a qualidade das repostas. Após a finalização dos testes o projeto teve a estrutura conforme a Figura 12.

Figura 12: Estrutura do projeto após a realização do teste



Fonte: O autor.

Pode-se observar que foi gerado um arquivo .csv chamado result.csv, no Quadro 3 é exibido um exemplo do seu conteúdo.

Quadro 3: result.csv

Respondeu	Similaridade maior	Similaridade resposta certa	Pergunta	Resposta certa	Resposta charbot
certo	0.796	0.796	fundamental a psicologia forense	estabelece modelos conceituais diferenciados dos utilizados na area clinica para que possam produzir conhecimento relevante do ponto de vista legal	necessidades a situação do treinamento estrategia que sera utilizada ou acao avaliacao dos resultados avaliacao e controle

Fonte: O autor.

No Quadro 3, observam-se principalmente as seguintes informações:

- Respondeu: pode receber dois valores certo ou errado e é a verificação se a alternativa que o *chatbot* escolheu com base na similaridade está certa ou errada;
- Similaridade maior: mostra a similaridade da resposta do *chatbot* e a resposta que tem a maior similaridade;
- Similaridade resposta certa: mostra a similaridade da resposta do *chatbot* com a resposta correta;
- pergunta: a pergunta que foi enviada para o *chatbot*;
- Resposta certa: a alternativa correta dentre as 5;
- Resposta chatbot: a resposta que o *chatbot* retornou;

Pode-se observar que mesmo o *chatbot* não respondendo exatamente igual à resposta certa ele conseguiu acertar a alternativa, isso acontece porque a resposta gerada teve maior semelhança com a resposta correta.

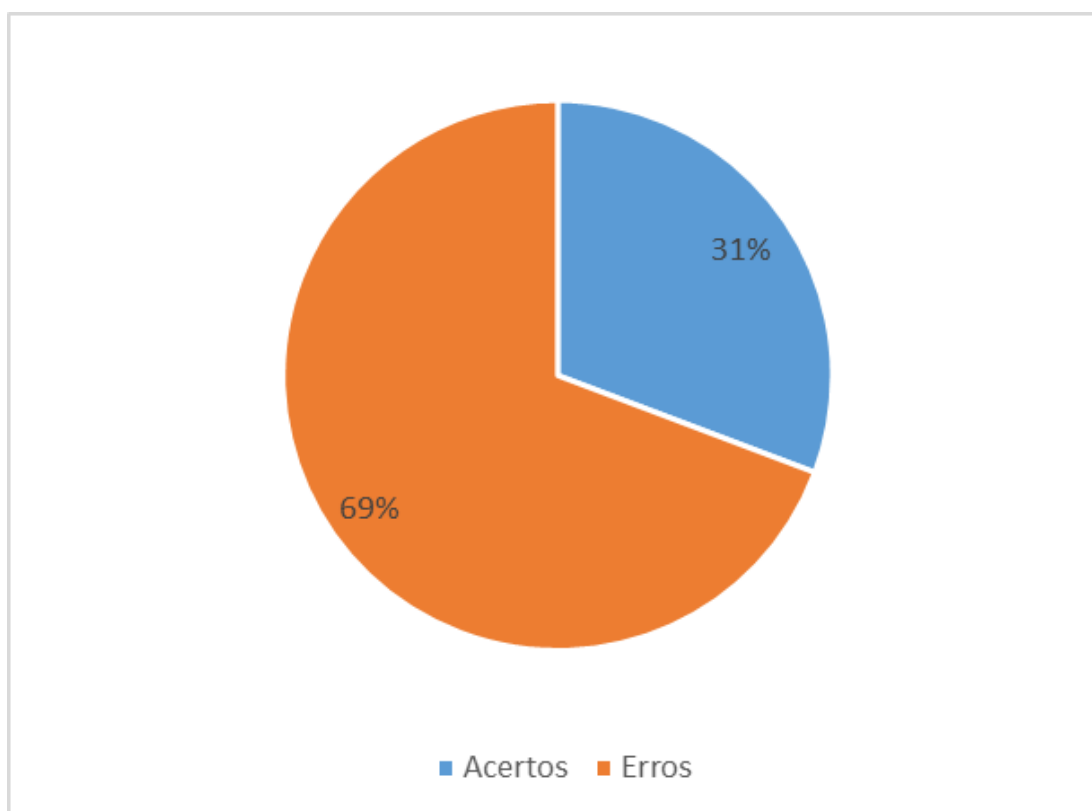
4 ANÁLISE DOS RESULTADOS

Com a conclusão da etapa de teste, passa-se para a análise das informações obtidas, os resultados são expostos em forma de gráficos para facilitar a didática das informações, onde é exibido 4 gráficos para cada matéria, podendo assim visualizar a porcentagem dos acertos e erros do *chatbot*, e a similaridade das repostas. A fim de facilitar o entendimento será adotado que 70% de acertos torna o *chatbot* eficaz, e valores inferiores o torna ineficazes.

4.1.1 Português

O *dataset* de português contém um total de 2440 questões, onde 1952 (80%) questões foram utilizadas para treinamento e 488 (20%) para testes. No Gráfico 2 pode ser observado a taxa de eficácia do *chatbot* na matéria de português, onde apresenta 31% de aproveitamento (acertos), ou seja, das 488 questões enviadas para o teste, o *chatbot* acertou 150 e errou 338 questões.

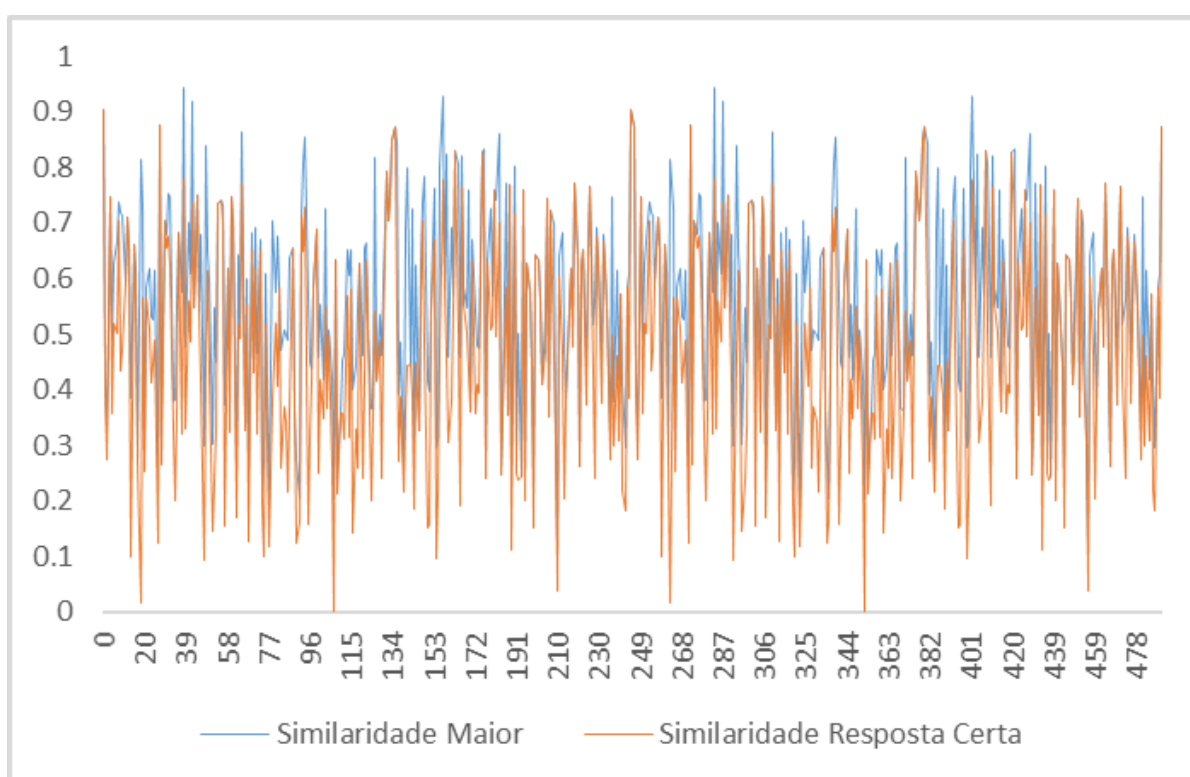
Gráfico 2: Acertos e erros (disciplina de português)



Fonte: O autor.

No Gráfico 3, é exibido 2 linhas, uma azul que representa a similaridade entre a resposta gerada pelo *chatbot* com a resposta que contém a maior similaridade entre as 5 alternativas. E a linha laranja representa a similaridade entre a resposta gerada pelo *chatbot* e a alternativa correta, quanto mais o valor se aproximar do número 1, mais similar são as respostas.

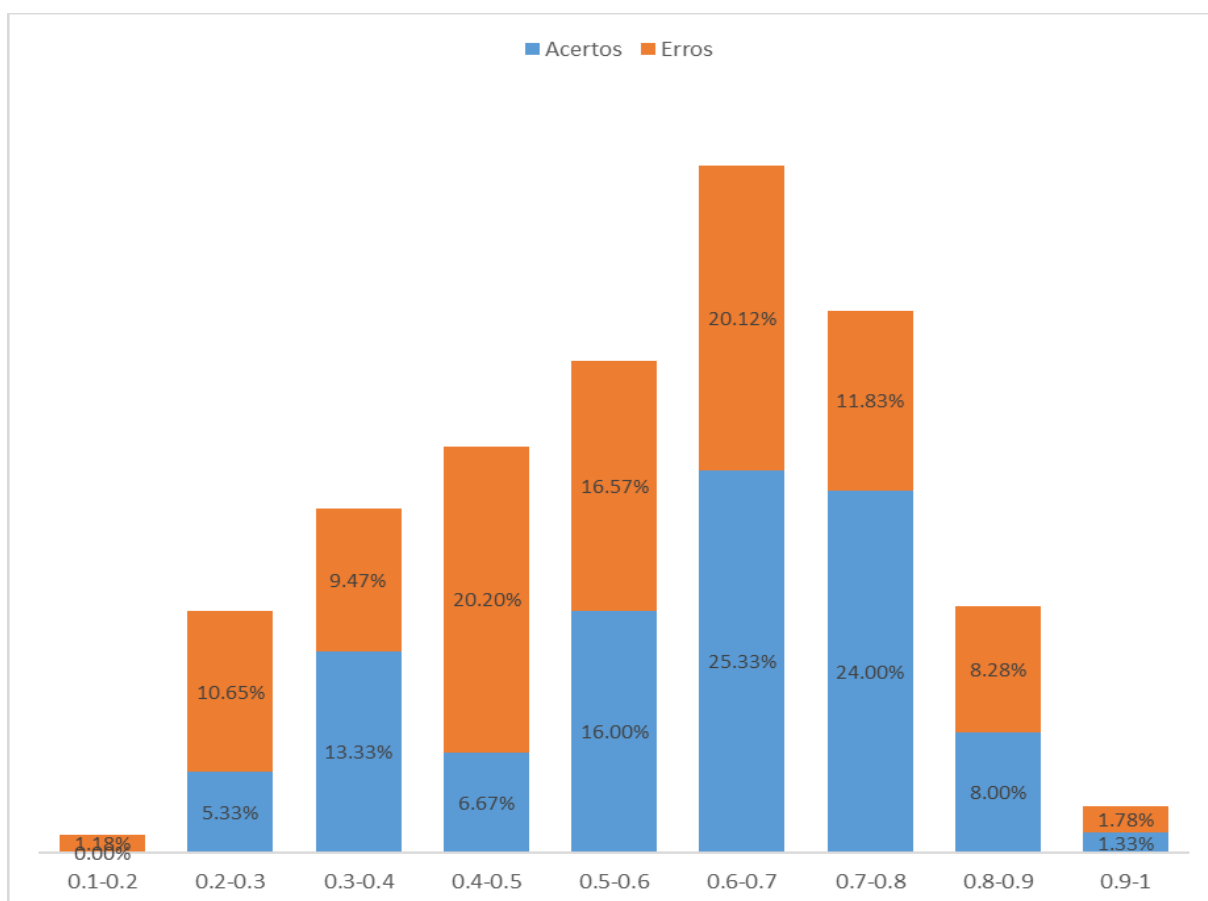
Gráfico 3: Similaridade resposta maior e similaridade resposta certa (disciplina de português)



Fonte: O autor.

O Gráfico 4 apresenta a frequência de ocorrências das similaridades entre a resposta do *chatbot* e a alternativa que tem maior similaridade (lembrando que tem 5 alternativas a, b, d, c, e e), onde é mostrado que a maior frequência de similaridade é entre 0.6 e 0.7, e teve média de 0.564. Isso significa que a maior parte das respostas geradas não são similares às alternativas escolhidas como correta pelo *chatbot*. As colunas em azul representam a porcentagem das respostas certas, enquanto a coluna de cor laranja representa a porcentagem das respostas erradas.

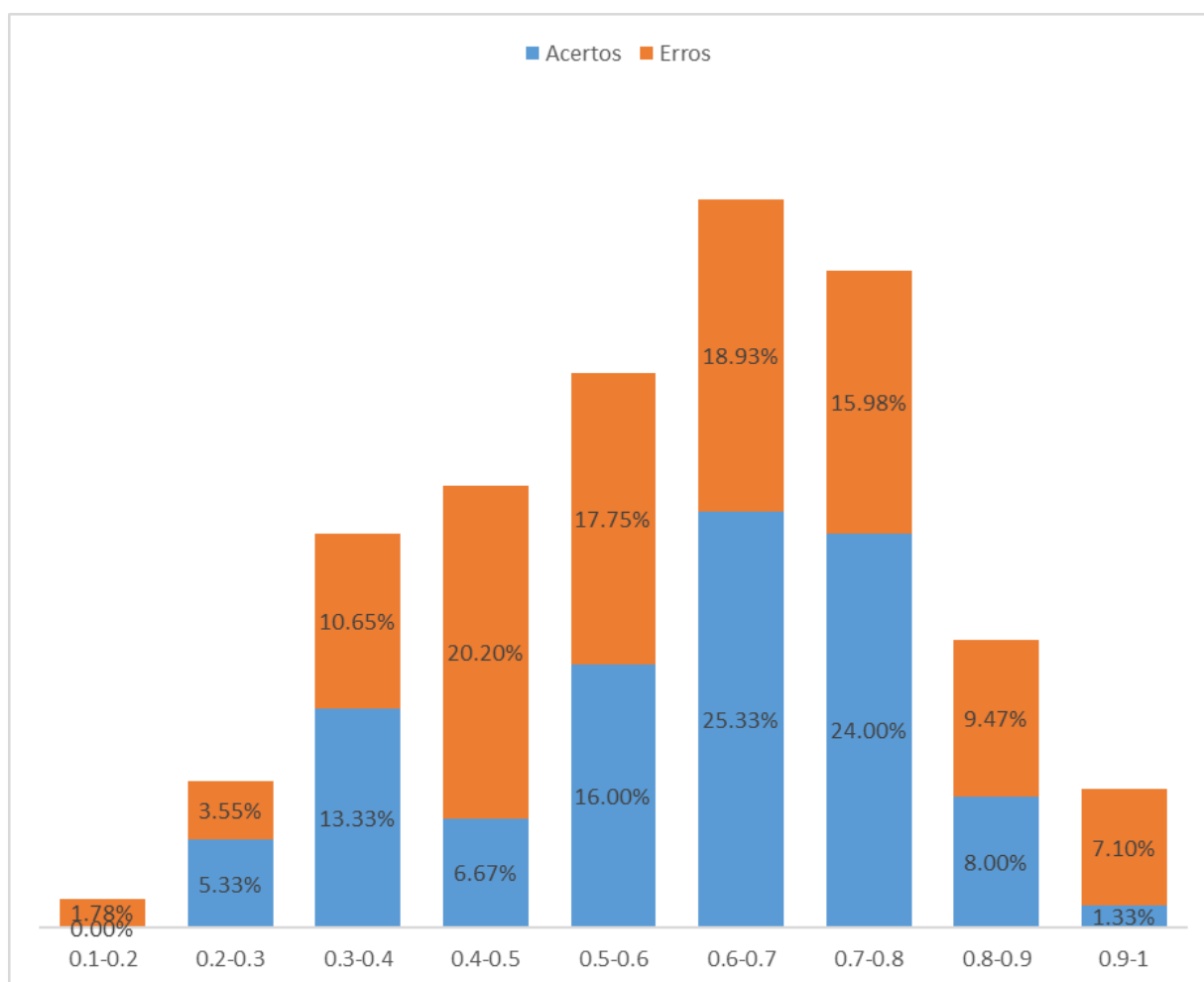
Gráfico 4: Frequência de ocorrências das similaridades entre a resposta do *chatbot* e a alternativa que tem maior similaridade (disciplina de português)



Fonte: O autor.

O Gráfico 5 apresenta a frequência de ocorrências das similaridades entre a resposta do chatbot e a alternativa correta, onde destaca-se com maior frequência o intervalo de 0.6 e 0.7. A maior parte das respostas estão abaixo de 0.7, sendo assim a maioria das respostas geradas pelo chatbot tem a média em 0.463. Ou seja, não tem muita similaridade com as alternativas corretas.

Gráfico 5: Frequência de ocorrências das similaridades entre a resposta do *chatbot* e a alternativa correta (disciplina de português)



Fonte: O autor.

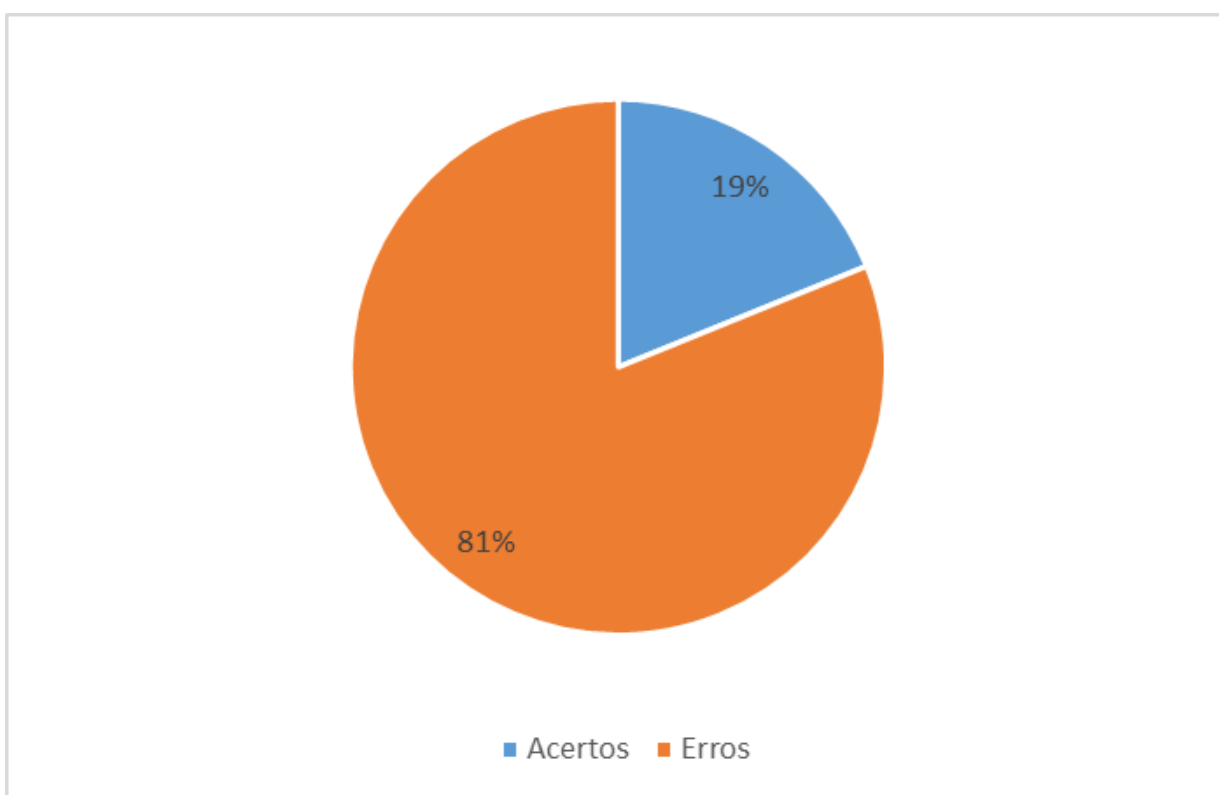
Como pode ser visto no

Gráfico 5 a porcentagem de acertos e a similaridade entre a resposta certa e a resposta do *chatbot* tiveram aproveitamento abaixo de 70%, sendo assim não seria eficaz para a matéria de português.

4.1.2 Medicina

O *dataset* medicina contém um total de 2175 questões, onde 1740 (80%) questões foram utilizadas para treinamento e 435 (20%) para teste. Pode ser visto no Gráfico 6 que o *chatbot* obteve um total de 19% de acertos e 81% de erros, sendo assim, essa matéria tem a menor taxa de acerto de comparados com a matéria anteriores.

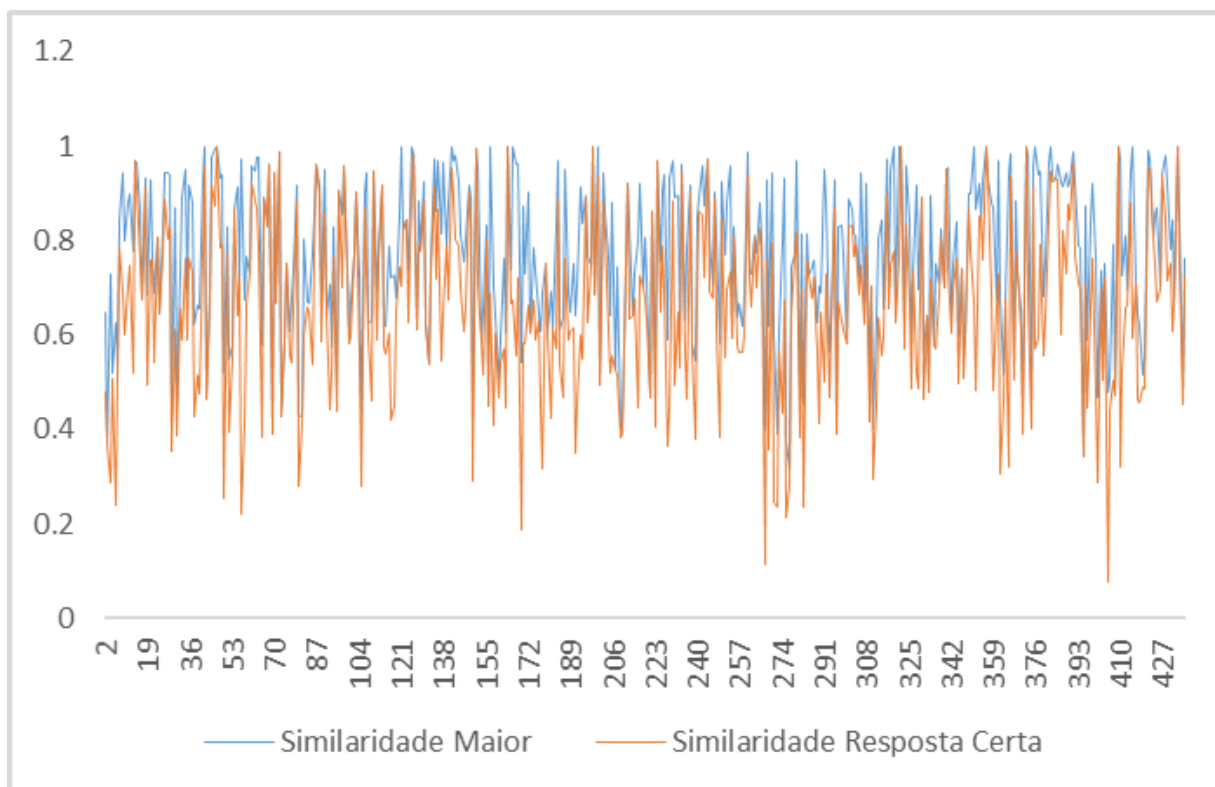
Gráfico 6: Acertos e erros (disciplina de medicina)



Fonte: O autor.

No Gráfico 7, pode-se ver que a similaridade foi mais próxima de 1 que nas matérias anteriores, em outras palavras a resposta do *chatbot* ficou mais similar às alternativas.

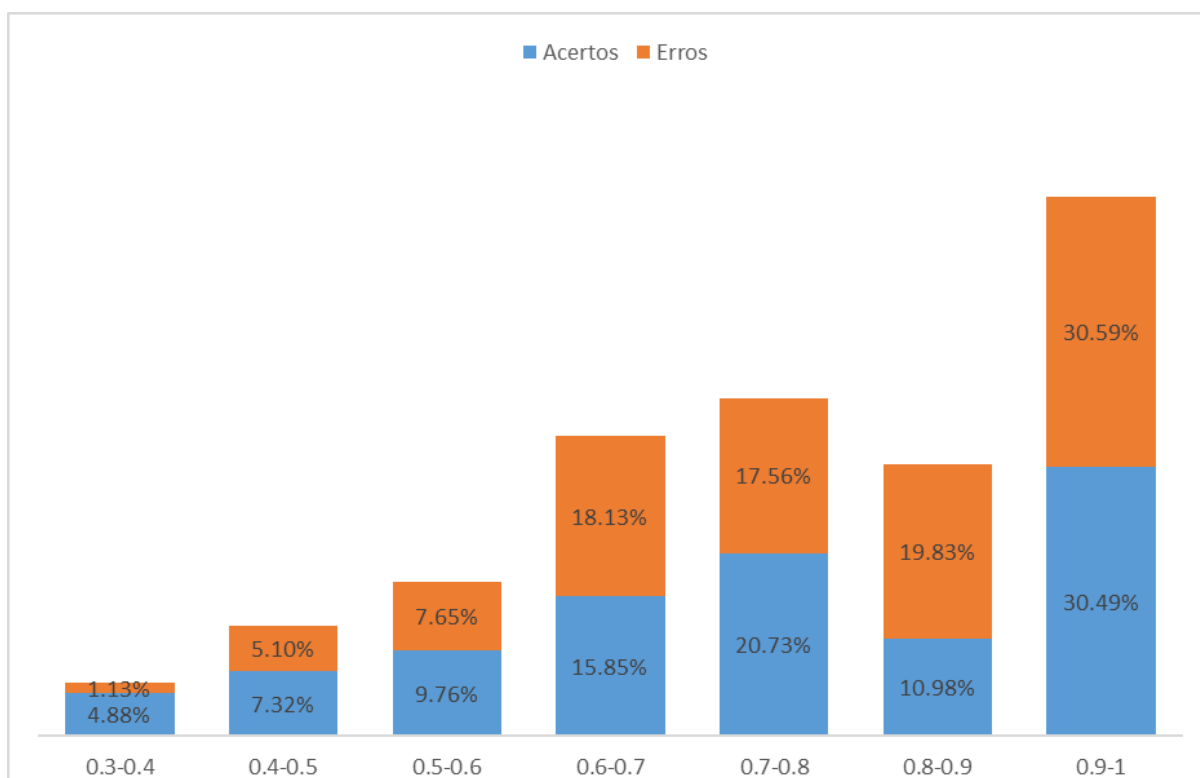
Gráfico 7: Similaridade resposta maior e similaridade resposta certa (disciplina de medicina)



Fonte: O autor.

O Gráfico 8 mostra que a maior frequência de similaridade ocorreu entre 0.9 e 1, ou seja, as alternativas que o *chatbot* marcou como certa, na maior parte das vezes foi maior que 0.9 de similaridade, e a média foi de 0.773, sendo assim, chegou mais próximo da alternativa que identificou como correta.

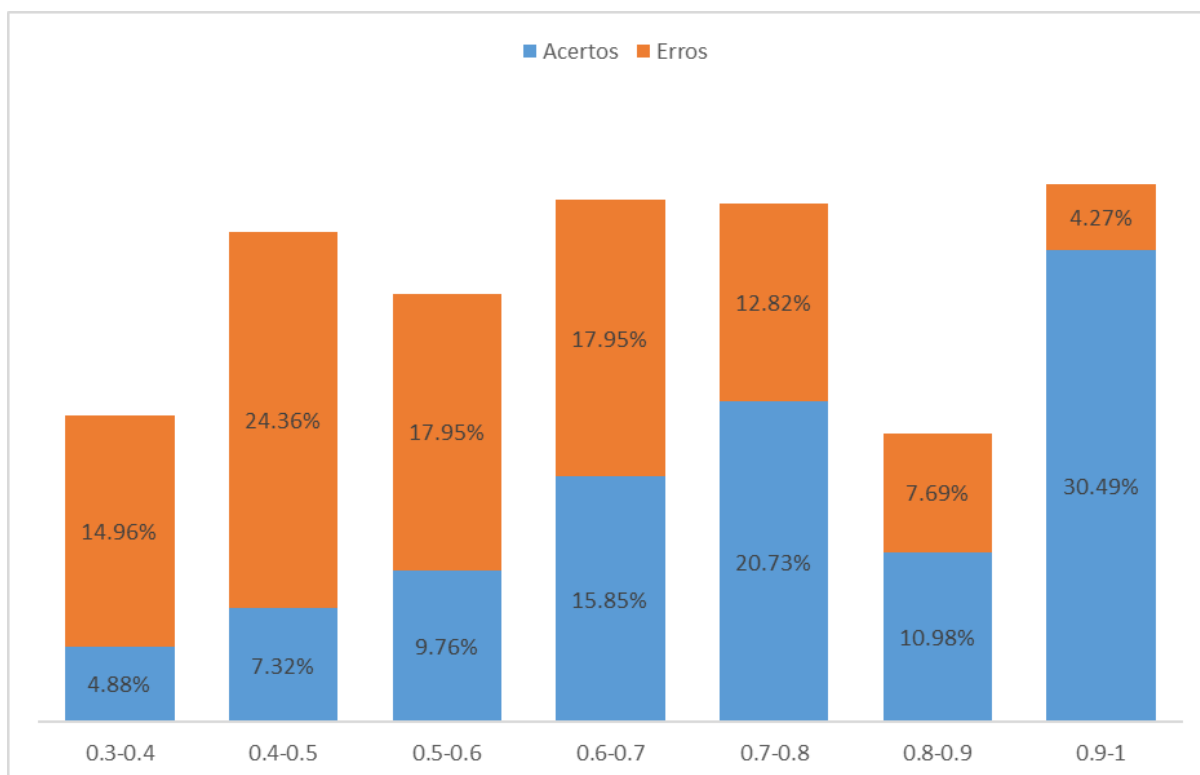
Gráfico 8: Frequência de ocorrências das similaridades entre a resposta do *chatbot* e a alternativa que tem maior similaridade (disciplina de medicina)



Fonte: O autor.

Já no Gráfico 9, pode ser visto que a maior similaridade foi entre 0.6 e 0.7, ou seja, não teve um aproveitamento maior que 70%, e uma média de 0.660.

Gráfico 9: Frequência de ocorrências das similaridades entre a resposta do *chatbot* e a alternativa correta (disciplina de medicina)



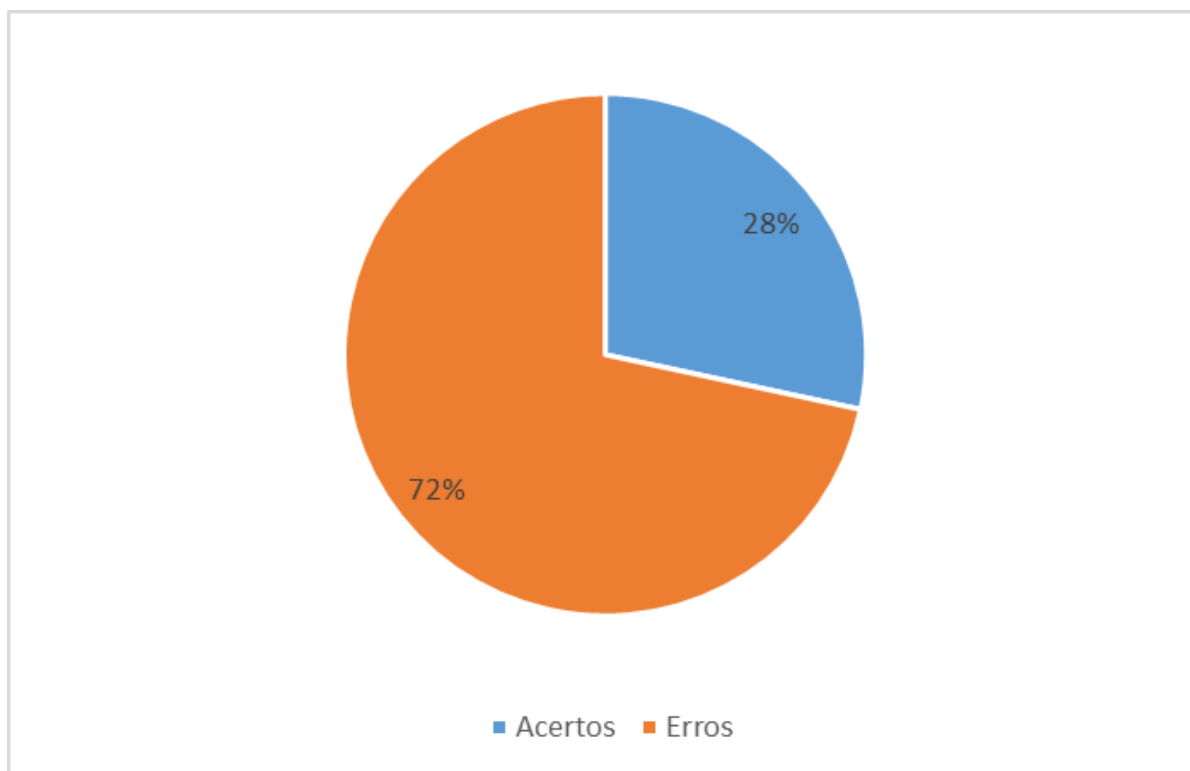
Fonte: O autor.

A disciplina de medicina saiu melhor entre a resposta do *chatbot* e a alternativa que tem maior similaridade, porém a taxa de acertos não foi superior a 70%, sendo assim considerado ineficaz.

4.1.1 Informática

O *dataset* de informática contém um total de 1745 questões, onde 1396 (80%) questões foram utilizadas para treinamento e 349 (20%) para teste.

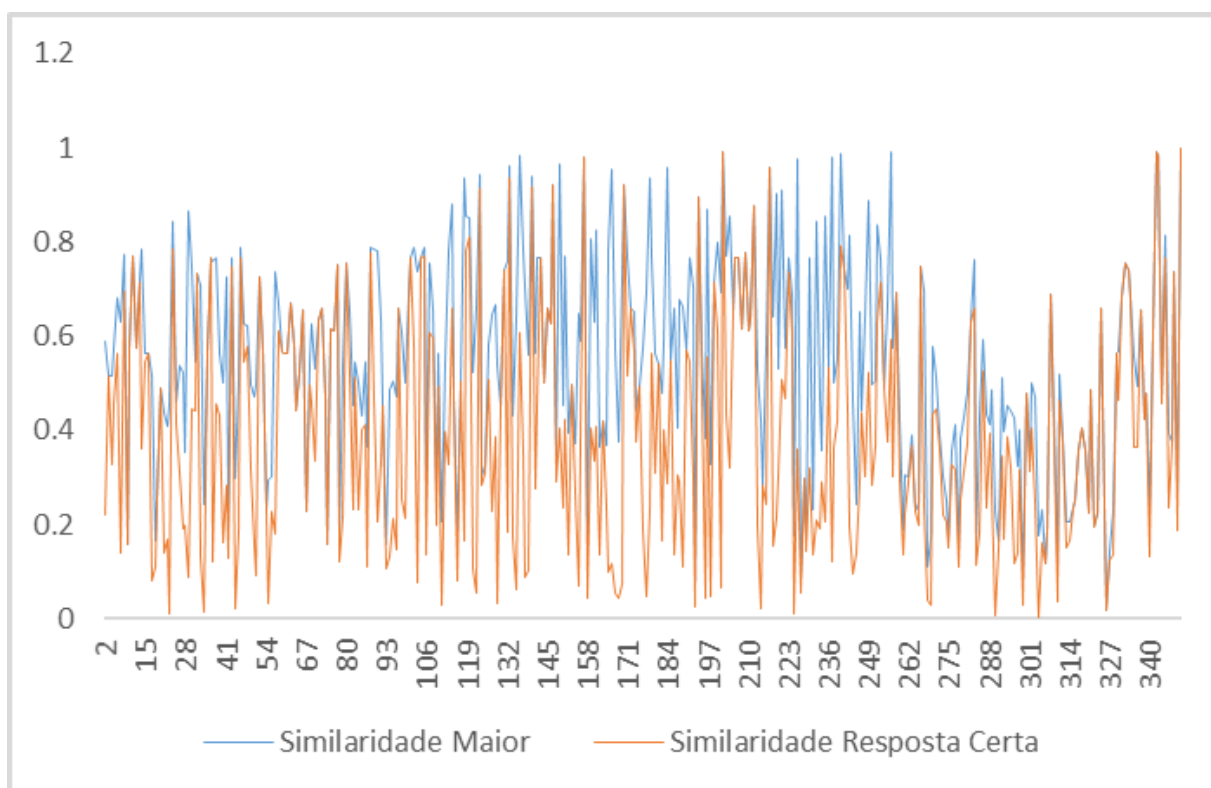
Gráfico 10: Acertos e erros (disciplina de informática)



Fonte: O autor.

No Gráfico 10 pode ser observado a taxa de eficácia do *chatbot* na matéria de informática, onde apresenta 28% de aproveitamento, ou seja, das 349 questões enviadas para o teste, o *chatbot* acertou 99 e errou 250 questões. Se comparado a matéria de português, esse teve uma piora em 3%.

Gráfico 11: Similaridade resposta maior e similaridade resposta certa (disciplina de informática)

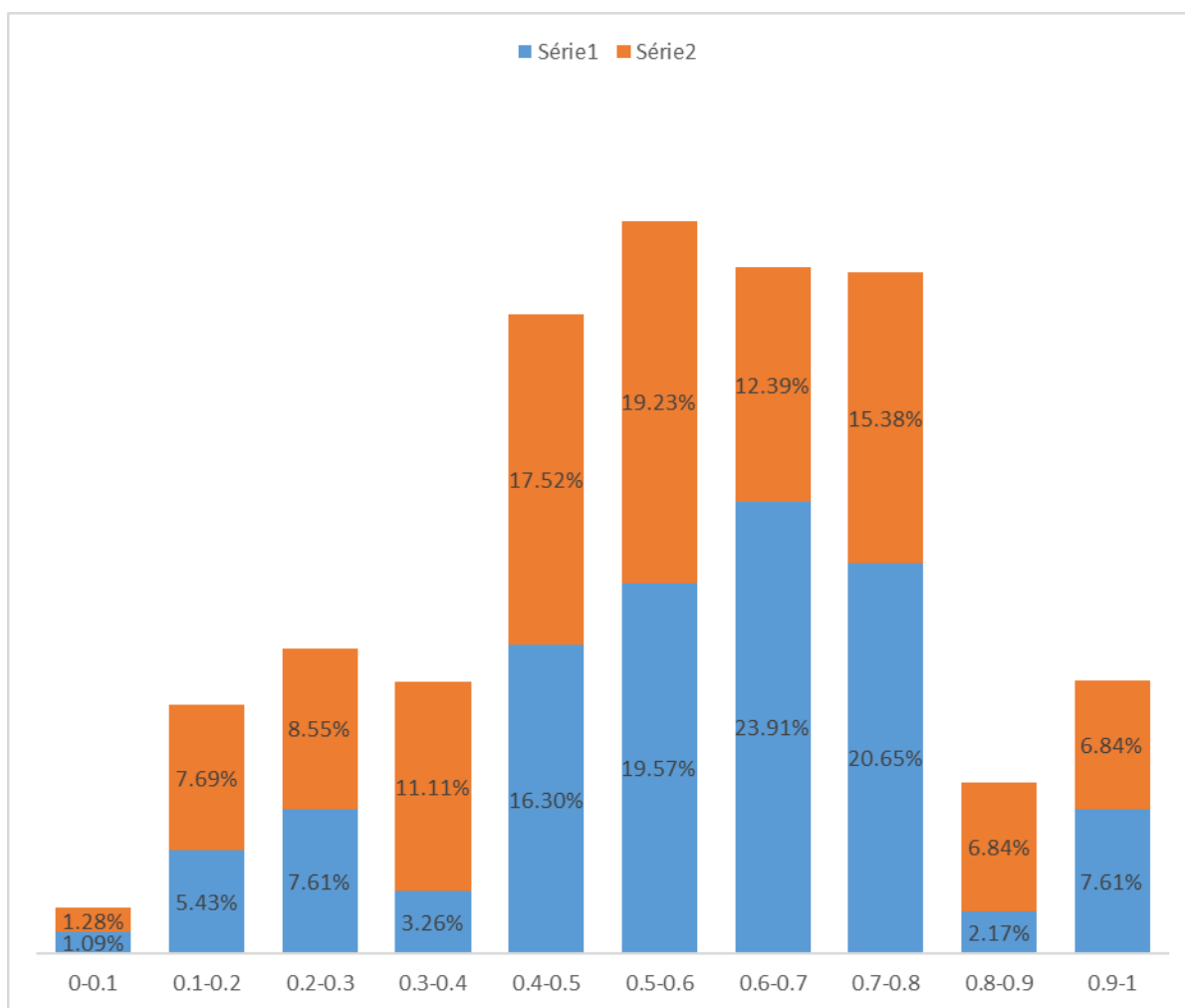


Fonte: O autor.

No

Gráfico 11, pode-se observar que teve uma variação maior de similaridade se comparado a matéria de português.

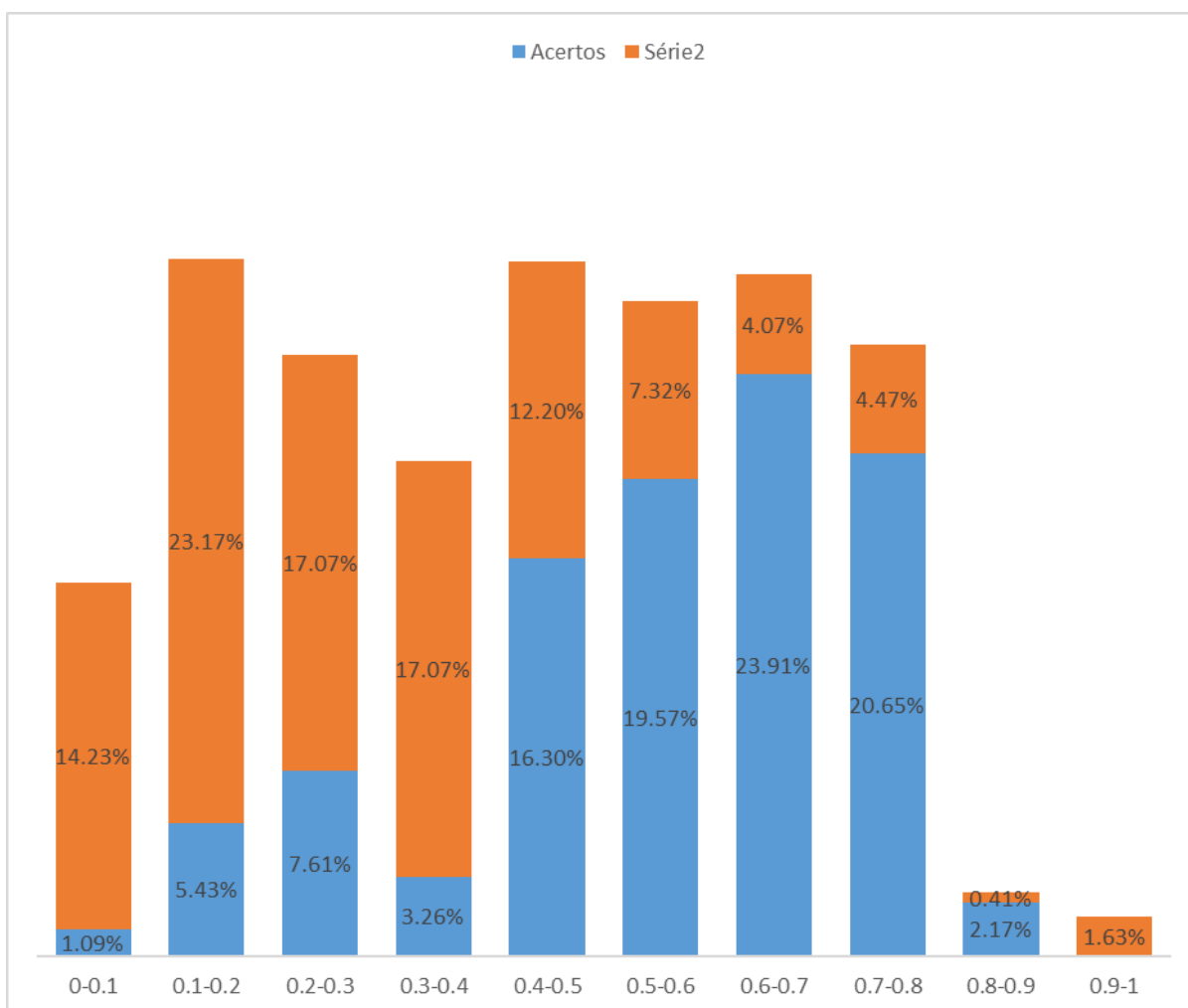
Gráfico 12: Frequência de ocorrências das similaridades entre a resposta do *chatbot* e a alternativa que tem maior similaridade (disciplina de informática)



Fonte: O autor.

O Gráfico 12 apresenta a frequência de ocorrências das similaridades entre a resposta do *chatbot* e a alternativa que tem maior similaridade, onde é mostrado que a maior frequência de similaridade é entre 0.5 e 0.6 e a média é 0.553, como na matéria de português.

Gráfico 13: Frequência de ocorrências das similaridades entre a resposta do *chatbot* e a alternativa correta (disciplina de informática)



Fonte: O autor.

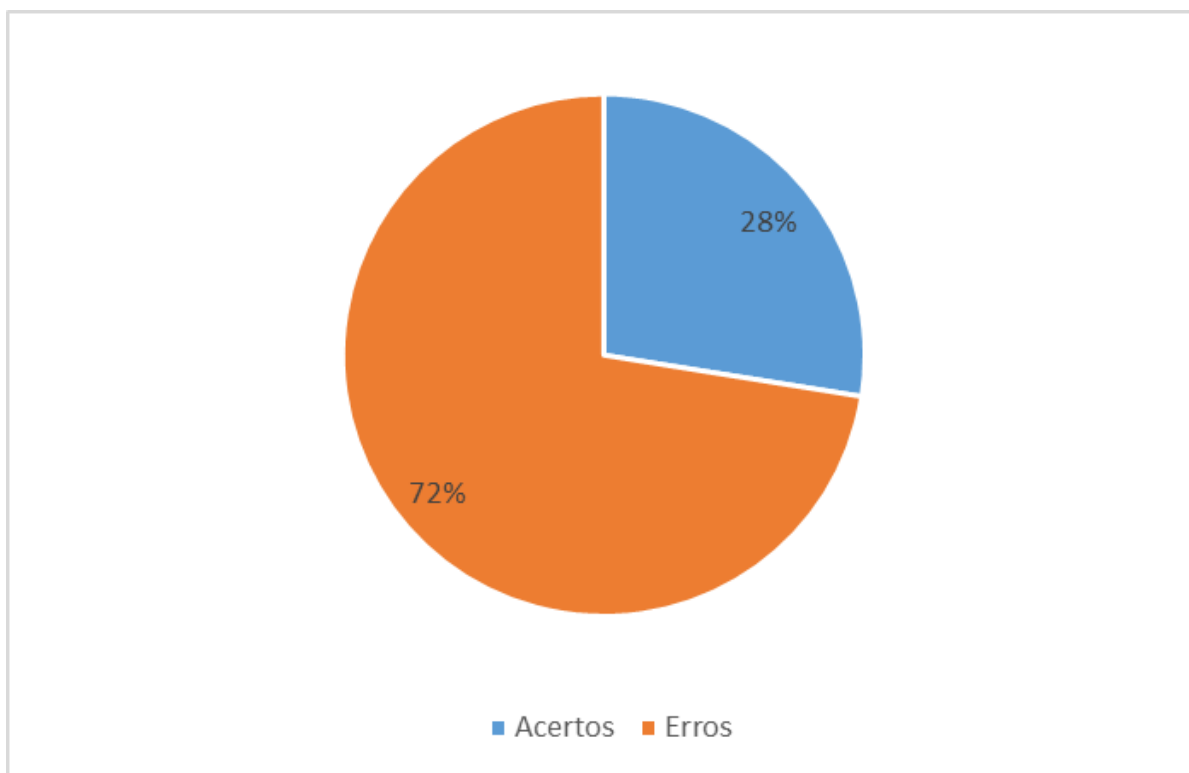
O Gráfico 13 apresenta a frequência de ocorrências das similaridades entre a resposta do *chatbot* e a alternativa correta, onde destaca-se com maior frequência o intervalo de 0.1 e 0.2 e média em 0.387, onde apresentou menor assertividade em relação a matéria de português.

A porcentagem de acertos e a similaridade entre a resposta certa e a resposta do *chatbot* tiveram aproveitamento abaixo de 70%, sendo assim não seria eficaz para a matéria de português.

4.1.2 Psicologia

O *dataset* psicologia contém um total de 2125 questões, onde 1700 (80%) questões foram utilizadas para treinamento e 425 (20%) para teste.

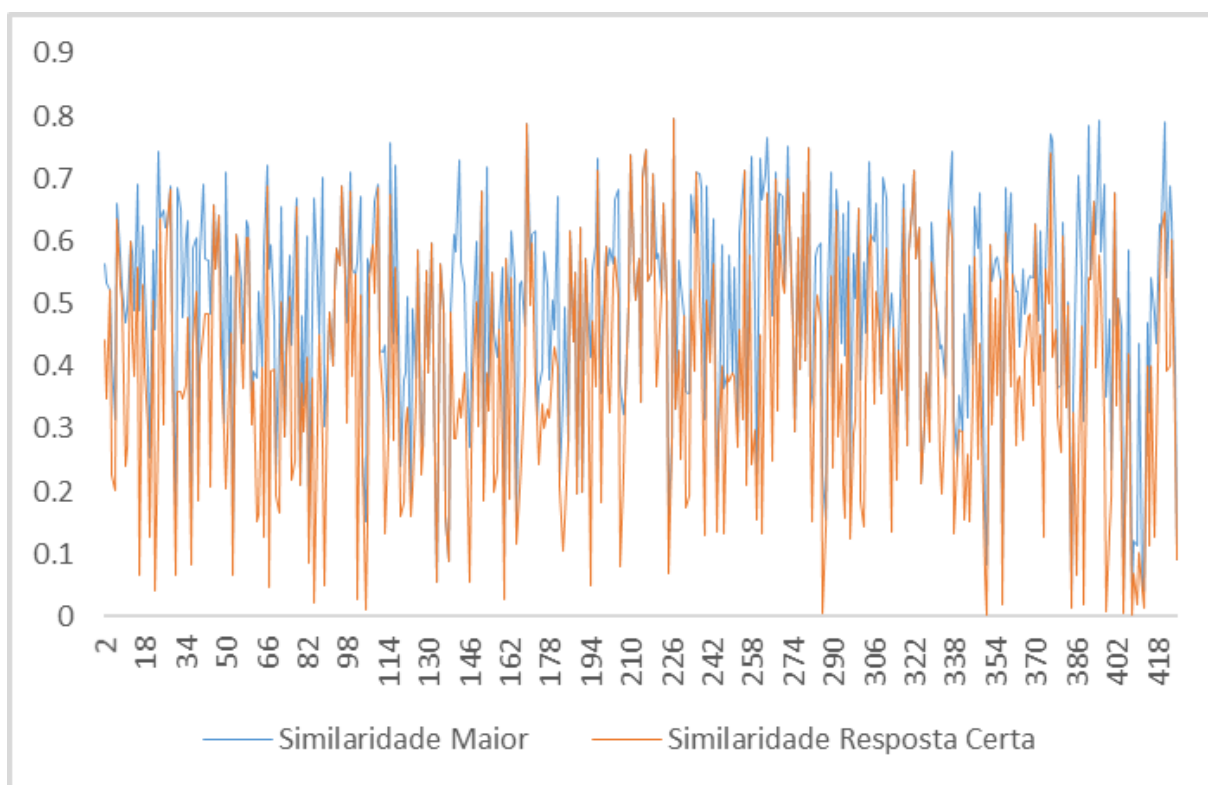
Gráfico 14: Acertos e erros (disciplina de medicina)



Fonte: O autor.

Pode-se observar no Gráfico 14 que o *chatbot* obteve um total de 28% de acertos e 72%, tendo assim o mesmo nível de acertos da disciplina de informática.

Gráfico 15: Similaridade resposta maior e similaridade resposta certa (disciplina de medicina)

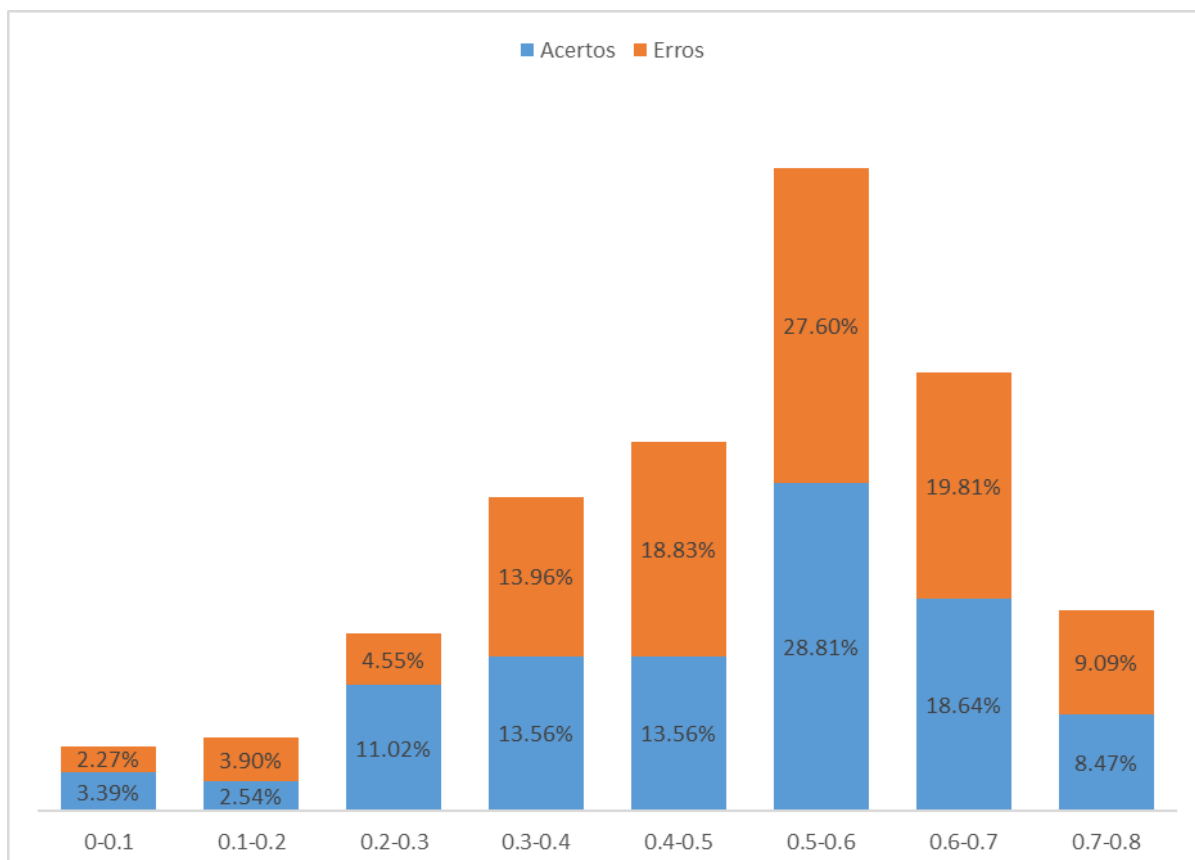


Fonte: O autor.

No Gráfico 15, pode-se observar que as variações das similaridades ficou mais uniforme se comparado ao

Gráfico 11, onde obtive a mesma porcentagem de acertos.

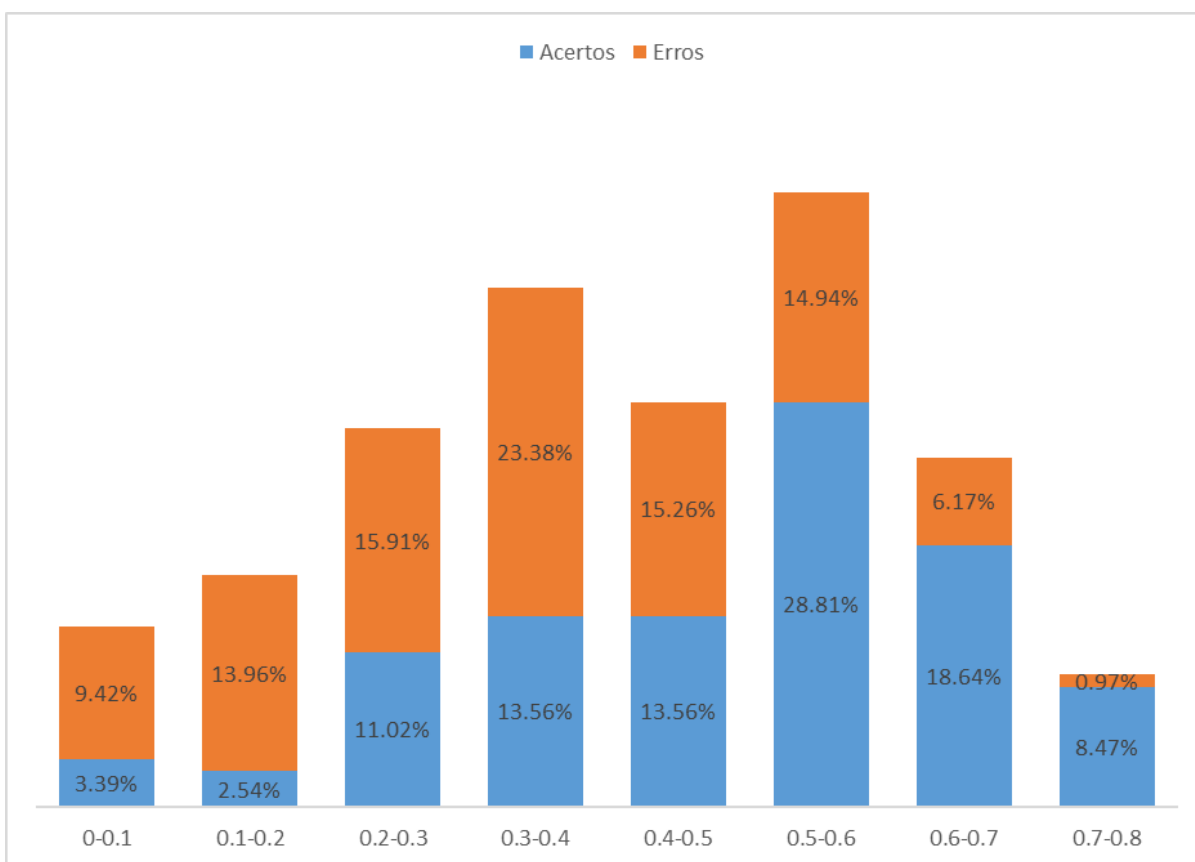
Gráfico 16: Frequência de ocorrências das similaridades entre a resposta do *chatbot* e a alternativa que tem maior similaridade (disciplina de medicina)



Fonte: O autor.

O Gráfico 16 mostra que a maior frequência de similaridade aconteceu entre 0.5 e 0.6 ficando assim com uma média de 0.498 que é abaixo da disciplina de informática por exemplo.

Gráfico 17: Frequência de ocorrências das similaridades entre a resposta do *chatbot* e a alternativa correta (disciplina de medicina)



Fonte: O autor.

Já no Gráfico 17 pode-se observar que a maior frequência de similaridade também está entre 0.5 e 0.6, porém com uma média de 0.382, e que também é menor se comparado com a disciplina de informática por exemplo.

Embora a porcentagem de acertos seja igual em ambas as disciplinas: informática e psicologia, o comportamento do *chatbot* se tornou melhor na matéria de informática, pois obteve uma maior média de similaridade entre as repostas.

4.1.3 Demais disciplinas

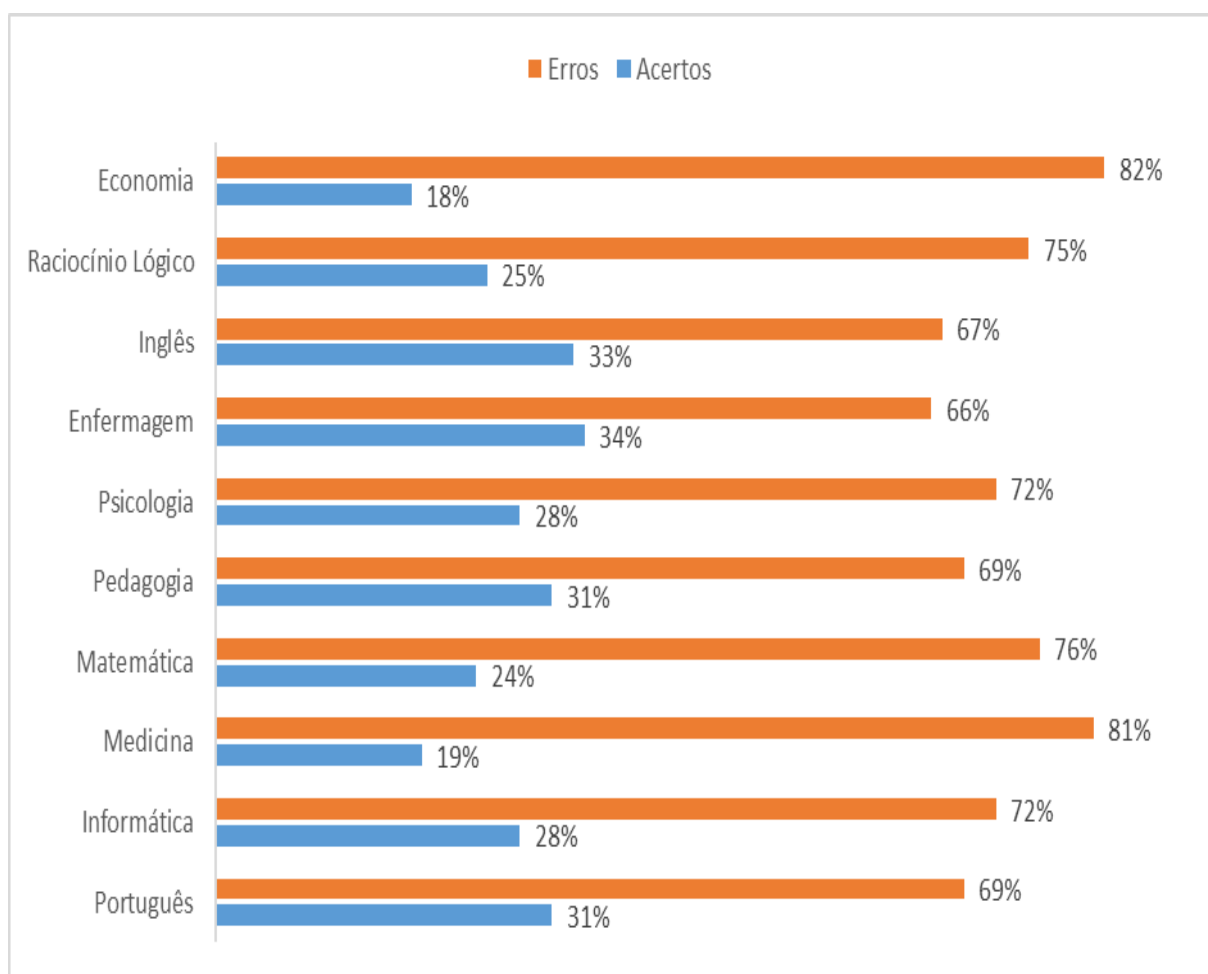
Os gráficos que representam os resultados das outras disciplinas podem ser vistos no APÊNDICE A, onde elas exibem a taxa de eficácia do *chatbot* com relação ao aproveitamento do mesmo, variação e frequência de ocorrência de similaridade quando comparado a outras disciplinas e alternativa que tem maior similaridade.

4.1.4 Análise geral das disciplinas

Nesta seção, passa-se a tratar de uma análise detalhada de resultados em relação às disciplinas, abordando especificidades.

No Gráfico 18, é mostrado a média de acertos e erros de todas as disciplinas treinadas e testadas pelo *chatbot*.

Gráfico 18: Acertos e erros por disciplinas



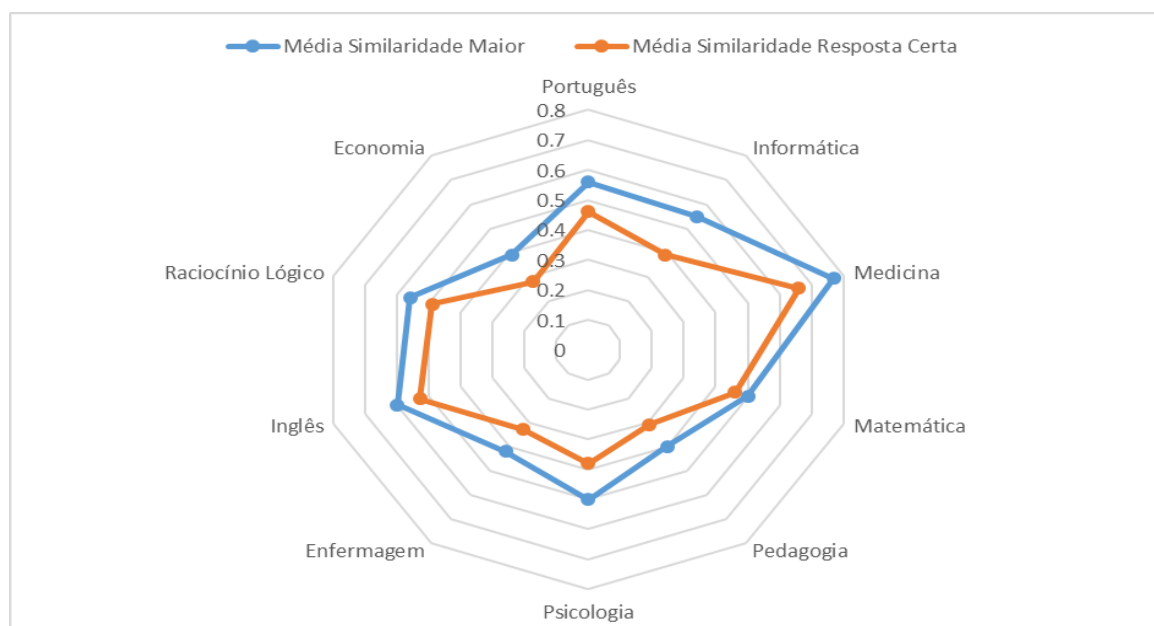
Fonte: O autor.

Pode-se observar no Gráfico 18 que o *chatbot* obteve uma melhor assertividade na matéria de enfermagem com média de 34%, e a pior na disciplina de economia onde obteve a média de 18%.

O Gráfico 19 é do tipo área, no qual contém duas linhas, sendo a de cor azul a representação de média de similaridade maior, ou seja, a alternativa que tem a maior similaridade com a resposta que o *chatbot* retornou, enquanto a linha laranja

representa a média de similaridade da resposta certa, isto é, a similaridade da resposta do *chatbot* em relação a alternativa correta.

Gráfico 19: similaridade por disciplinas



Fonte: O autor.

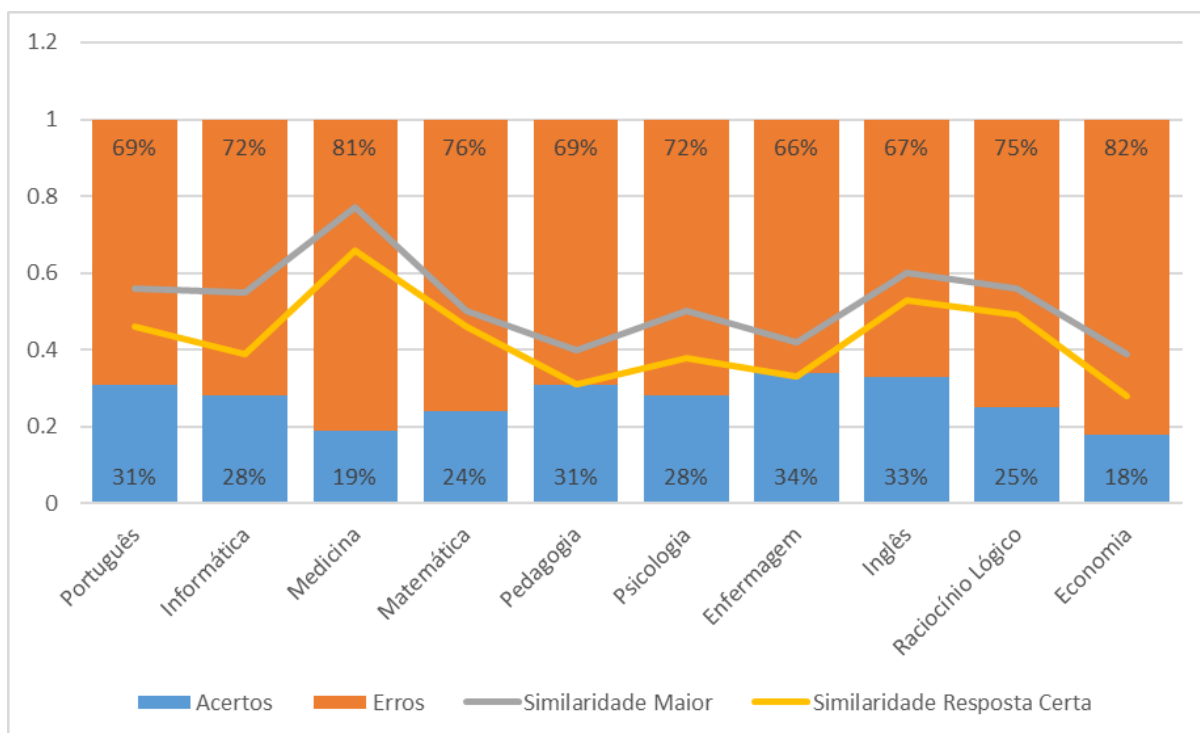
No Gráfico 19, pode-se observar que, quando o *chatbot* foi testado na disciplina de medicina, o mesmo alcançou a média de similaridade mais alta, tanto na maior similaridade entre as alternativas (linha azul), quanto na similaridade da resposta do chatbot com alternativa correta (linha laranja), sendo que as medias foram:

- similaridades entre a resposta do *chatbot* e a alternativa que tem maior similaridade = 0.77
- similaridades entre a resposta do *chatbot* e a alternativa correta = 0.66

A média geral de acertos das disciplinas foi de 27% enquanto a média de erros foi de 73%.

Já o Gráfico 20 é do tipo barra com linhas, onde as barras representam a porcentagem de acertos e erros por disciplinas, enquanto as linhas representam as similaridades das disciplinas, sendo que a de cor cinza é entre a resposta do *chatbot* e a alternativa de maior similaridade, e a de cor amarela é entre a resposta do *chatbot* e a alternativa correta.

Gráfico 20: Acertos e erros com similaridades



Fonte: O autor.

Pode-se observar no Gráfico 20 que a similaridade não teve um padrão seguindo a taxa de acertos e erros, sendo que na matéria de medicina, onde teve a maior similaridade, e a taxa de acertos foi de somente 19%, enquanto enfermagem que obteve 34% de acertos teve similaridade média de:

- similaridades entre a resposta do *chatbot* e a alternativa que tem maior similaridade = 0.41
- similaridades entre a resposta do *chatbot* e a alternativa correta = 0.33

Sendo assim, a matéria com menor porcentagem de erros, teve também a menor taxa de similaridade, tanto entre a alternativa de maior similaridade quanto a alternativa correta.

5 CONCLUSÃO

Como foi citado no início dessa pesquisa as avaliações escolares tem como objetivo obter informações sobre o rendimento dos alunos no processo de ensino-aprendizagem. Esse estudo se iniciou com a concepção de analisar o *chatbot* desenvolvido com o modelo *seq2seq* utilizando a biblioteca *TensorFlow*, afim de identificar se o mesmo é ou não eficaz para atuação na área escolar.

Os resultados obtidos pelo treinamento do modelo de 10.000 épocas foram coletados e analisados em forma de gráficos utilizando a ferramenta Excel,

É importante ressaltar que as médias de acertos por disciplinas foram menores que 70%, sendo que a maior delas atingiu 34%. Sendo assim, pode-se fazer testes com diferentes modelos de redes neurais com maior quantidade de questões, a fim de alcançar resultados mais satisfatórios, tanto quanto a similaridade das repostas quanto a assertividade do chatbot.

Notoriamente o presente trabalho abre novas possibilidades para explorar o campo de IA (Inteligência Artificial) no meio acadêmico, é evidenciado que mesmo não atingindo uma boa assertividade, o chatbot conseguiu responder uma certa porcentagem das respostas de forma correta, sendo que a média em porcentagem total em todas as disciplinas de acertos foi de 27% enquanto a média de erros foi de 73%, sendo assim, o chatbot desenvolvido utilizando o modelo *seq2seq* do *TensorFlow*, teve um aproveitamento inferior a 70% de assertividade.

6 TRABALHO FUTURO

Como possível trabalho futuro, pode-se apontar:

- Desenvolvimento da rede neural do chatbot utilizando o modelo seq2seq com um mecanismo de atenção ao modelo de sequência a sequência, e realização do treinamento com maior quantidade de dados, com classificação de disciplinas e assuntos das questões, afim de obter uma melhor eficácia do chatbot.

REFERÊNCIAS

BOUTIN, Paul. *Um bot precisa de processamento de linguagem natural?* 2017. Disponível em: <<https://chatbotsmagazine.com/does-a-bot-need-natural-language-processing-c2f76ab7ef11>>. Acesso em: 07 de novembro de 2019.

CALDEIRA, Anna Maria Salgueiro. *Avaliação e processo de ensino-aprendizagem*. Belo Horizonte: Presença Pedagógica, 2000. v. 3, p. 122.

CAVALHEIRO, Juliana. *10 ferramentas / Frameworks de código aberto para IA*. Disponível em: <<https://blog.estabil.is/10-ferramentas-frameworks-de-codigo-aberto-para-ia>>. Acesso em: 16 de novembro de 2019.

COPELAND, Brian Jack *Inteligência artificial*. 2018. Disponível em: <<https://www.britannica.com/technology/artificial-intelligence>>. Acesso em: 07 de novembro de 2019.

DACOMBE, James. *An introduction to Artificial Neural Networks (with example)*. 2017. Disponível em: <<https://medium.com/@jamesdacombe/an-introduction-to-artificial-neural-networks-with-example-ad459bb6941b>>. Acesso em: 08/12/2019.

DAVYDOVA, Olga. *7 tipos de redes neurais artificiais para processamento de linguagem natural*. 2017. Disponível em: <<https://medium.com/@datamonsters/artificial-neural-networks-for-natural-language-processing-part-1-64ca9ebfa3b2>>. Acesso em: 21 de junho de 2019.

HAYKIN, Simon. *Neural Networks: A Comprehensive Foundation*. 2. ed. Upper Saddle River, NJ, USA: Prentice Hall, 1998. 823p

HOCHREITER, Sepp. e SCHMIDHUBER, Jürgen. *Long Short-Term Memory*. 8. ed, Massachusetts: Neural Comput. v.09, p. 1735–1780, nov. 2006. Disponível em: <<http://dx.doi.org/10.1162/neco.1997.9.8.1735>>. Acesso em: 06 de novembro de 2019.

INFOPÉDIA, *chat in Dicionário da Língua Portuguesa*. Porto Seguro: Porto Editora, 2003. Disponível em: <<https://www.infopedia.pt/dicionarios/lingua-portuguesa/chat>>. Acesso em: 06 de novembro de 2019.

JEANS, Nathalie. *How I Classified Images With Recurrent Neural Networks*. 2013. Disponível em: <<https://medium.com/@nathaliejeans7/how-i-classified-images-with-recurrent-neural-networks-28eb4b57fc79>>. Acesso em: 07/11/2019.

LAVEN, Simon. *The Simon Lavel Homepage*. 2006. Disponível em: <<http://www.simonlaven.com/>>. Acesso em: 26 de maio 2019.

LI, Susan. *Machine Learning for Text Classification Using SpaCy in Python*. 2018. Disponível em: <<https://towardsdatascience.com/machine-learning-for-text-classification-using-spacy-in-python-b276b4051a49>>. Acesso em: 10/11/2019.

LIMA, Danielle. Quais os tipos de chatbots? 2018. Disponível em: <<https://medium.com/disrupcaocognitiva/quais-os-tipos-de-chatbots-9268b2888d18>>. Acesso em: 15 de junho de 2019.

LOUREIRO, Rodrigo. *Entenda de uma vez por todas o que é um bot e como ele funciona*. 2016. Disponível em: <<https://olhardigital.com.br/noticia/entenda-de-uma-vez-por-todas-o-que-e-um-bot-e-como-ele-funciona/57075>>. Acesso em: 9 de junho de 2019.

LUCKESI, Cipriano Carlos. *Avaliação da aprendizagem escolar: estudos e proposições*. 17. ed. São Paulo: Cortez, 2005. 47p.

LUCKESI, Cipriano Carlos. *Avaliação da aprendizagem na escola: reelaborando conceitos e recriando a prática*. 1. ed. Salvador: Malabares Comunicação e Eventos, 2003.

MARR, Bernard. *What Are Artificial Neural Networks - A Simple Explanation For Absolutely Anyon*. 2018. Disponível em: <<https://www.forbes.com/sites/bernardmarr/2018/09/24/what-are-artificial-neural-networks-a-simple-explanation-for-absolutely-anyone/#449803631245>>. Acesso em: 08/12/2019.

MENEZES, Ebenezer Takuno. e SANTOS, T. H. *Verbete avaliação escolar. Dicionário Interativo da Educação Brasileira - Educabrazil*. São Paulo: Midiamix, 2001. Disponível em: <<https://www.educabrazil.com.br/avaliacao-escolar/>>. Acesso em: 06 de novembro de 2019.

MILIOZZI, João. *Chatbots: conheça a história dessa fascinante tecnologia*. 2017. Disponível em: <<https://ibramerc.liveuniversity.com/2017/12/15/chatbots-e-a-historia-dessa-fascinante-tecnologia/>>. Acesso em: 07/11/2019.

NUNES, Teresa. *O que é aprendizagem?* 2017. Disponível em: <<https://pontodidatica.com.br/o-que-e-aprendizagem/>>. Acesso em: 08/12/2019.

OLIVEIRA, Lucas Reis. *Você conhece os Chatbots? Descubra aqui o que são e como usá-los para otimizar o atendimento ao cliente.* 2019. Disponível em: <<https://rockcontent.com/blog/chatbots/>>. Acesso em: 07/11/2019.

ROCHA, Hugo. *Chatbot: o que é, para que serve, como funciona e como criar o seu.* 2018. Disponível em: <<https://klickpages.com.br/blog/chatbot-o-que-e/>>. Acesso em: 26 de maio de 2019.

ROUSE, Margaret. *AI (inteligência artificial).* 2018. Disponível em: <<https://searchenterpriseai.techtarget.com/definition/AI-Artificial-Intelligence>>. Acesso em: 07 de novembro de 2019.

SAS. *Processamento de Linguagem Natural (PNL).* 2019. Disponível em: <https://www.sas.com/en_us/insights/analytics/what-is-natural-language-processing-nlp.html>. Acesso em: 16 de junho de 2019.

SCHAPPO, Vanessa. *Chatbot: o que é, quais são as suas vantagens e como usar na sua empresa.* 2018. Disponível em: <<https://resultadosdigitais.com.br/blog/o-que-e-chatbot/>>. Acesso em: 06 de novembro de 2019.

SKINNER, Burrhus Frederic. *Tecnologia do ensino.* São Paulo: Epu. 1972. 164p.

SOUZA, Amanda. *Chatbots na educação: saiba como utilizá-lo no processo de ensino e aprendizagem.* 2018. Disponível em: <<http://blog.ellodeideias.com.br/author/amandasouza/>>. Acesso em: 07 de novembro de 2019.

TAFNER, Malcon Anderson. *What Are Artificial Neural Networks.* 1998. Disponível em: <<http://www.cerebromente.org.br/n06/tecnologia/rna.htm>>. Acesso em: 07 de novembro de 2019.

TEIXEIRA, João de Fernandes. *Robots, intencionalidade e inteligência artificial.* São Paulo: Brasiliense. v. 14, p. 109-121, 1991. Disponível em: <<http://www.scielo.br/pdf/trans/v14/v14a06.pdf>>. Acesso em: 07 de novembro de 2019.

TURING, Alan Mathison. *Computing Machinery and Intelligence*. Mind: New Series, v.59, n.236, p. 433-460, oct. 1950. Disponível em: <https://www.jstor.org/stable/2251299?origin=JSTOR-pdf&seq=1#page_scan_tab_contents>. Acesso em: 07 de novembro de 2019.

UNRUH, Amy. *What is the TensorFlow machine intelligence platform?*. 2017. Disponível em: <<https://opensource.com/article/17/11/intro-tensorflow>>. Acesso em: 01/11/2019.

VIGOTSKI, Lev. *A construção do pensamento e da linguagem*. São Paulo: Martins Fontes, 2000.

WAY, New *Chatbots na educação: transformando ensino e engajamento*. 2017. Disponível em: <<https://www.gruponewway.com.br/2017/06/22/chatbots-na-educacao-transformando-ensino-e-engajamento/>>. Acesso em: 18 de junho de 2019.

WEIZENBAUM, Joseph. *Eliza, a computer program for the study of natural language communication between man and machine*. Massachusetts, Communications of the ACM, v.09, n.01, p.36-45, jan. 1966. Disponível em: <http://www.universelle-automation.de/1966_Boston.pdf>. Acesso em: 07 de novembro de 2019.

YEGULALP, Serdar. *What is TensorFlow? The machine learning library explained*. 2019. Disponível em: <<https://www.infoworld.com/article/3278008/what-is-tensorflow-the-machine-learning-library-explained.html>>. Acesso em: 01/11/2019.

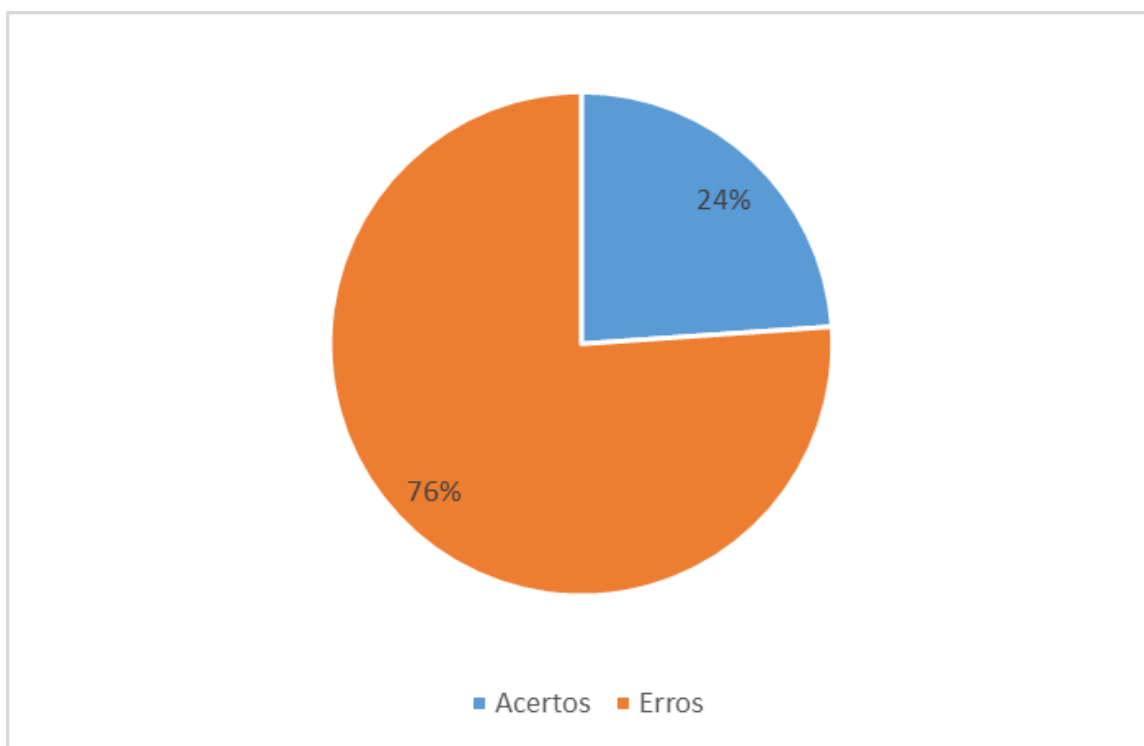
APÊNDICE A

Os gráficos são referentes às disciplinas de matemática, pedagogia, enfermagem, inglês, raciocínio lógico e economia, especificamente, sendo quatro gráficos para cada disciplina.

I. Matemática

O dataset de matemática contém um total de 2655 questões, onde 2124 (80%) questões foram utilizadas para treinamento e 531 (20%) para teste. Pode-se observar que na **Erro! Fonte de referência não encontrada.**¹ que o chatbot obteve um total de 24% acertos e 76% erros. Teve um desempenho melhor que a disciplina de medicina, porém inferior a português e informática.

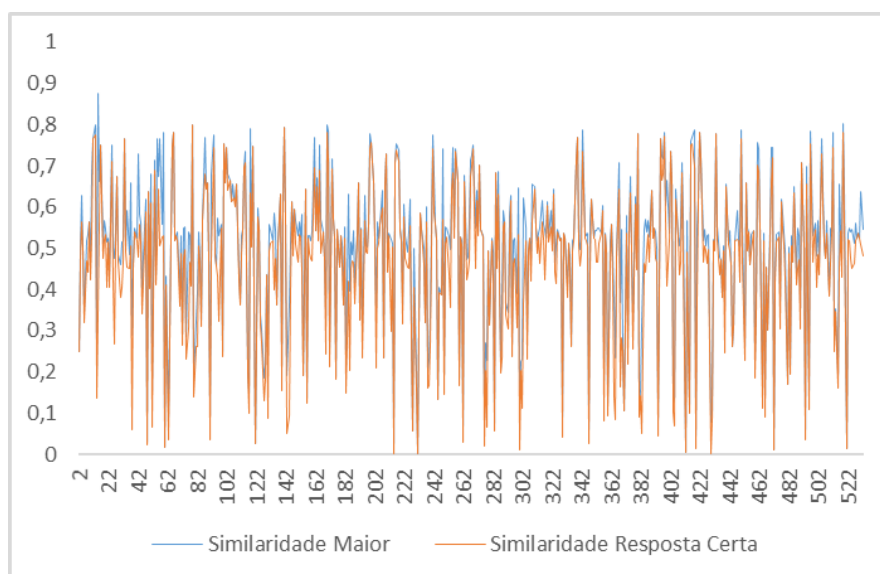
Gráfico 1: Acertos versus erros (disciplina de matemática)



Fonte: O autor.

No Gráfico 2, pode-se observar que a linha não chegou a atingir o número um em nenhum momento, isso significa que na disciplina de matemática o chatbot não conseguiu ter uma alternativa 100% similar em nenhum momento.

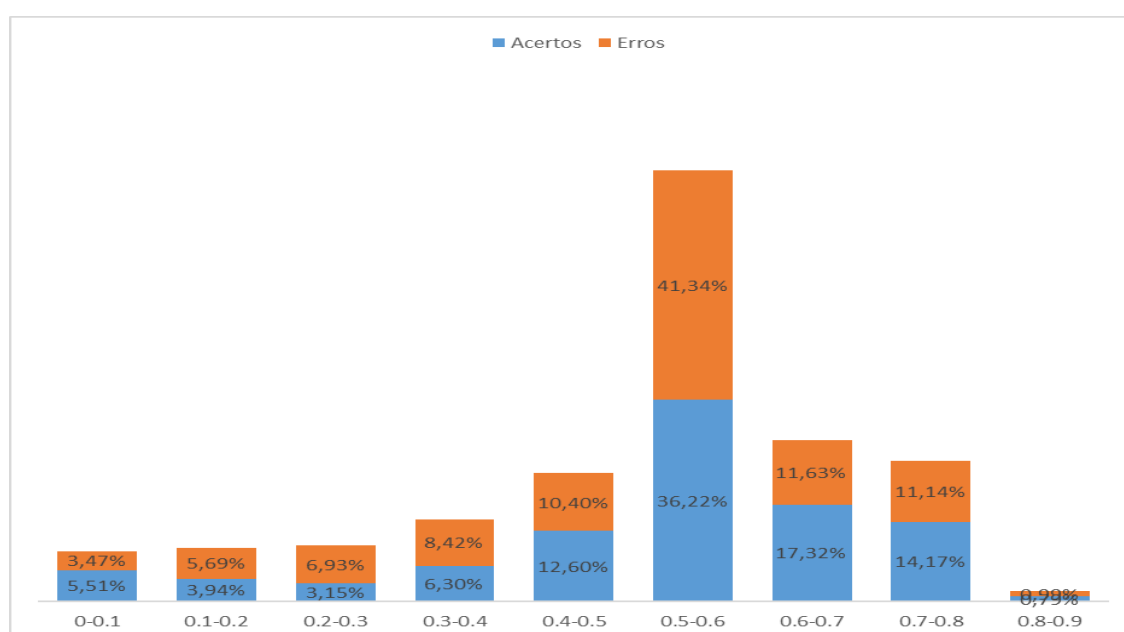
Gráfico 2: Similaridade resposta maior versus similaridade resposta certa (disciplina de matemática)



Fonte: O autor.

No Gráfico 3 pode-se observar que a maior frequência de similaridade ocorreu entre 0.5 e 0.6, o que significa que ao escolher uma alternativa e destaca-la como correta, na maioria das vezes levou em consideração uma média de similaridade de 0.502256425.

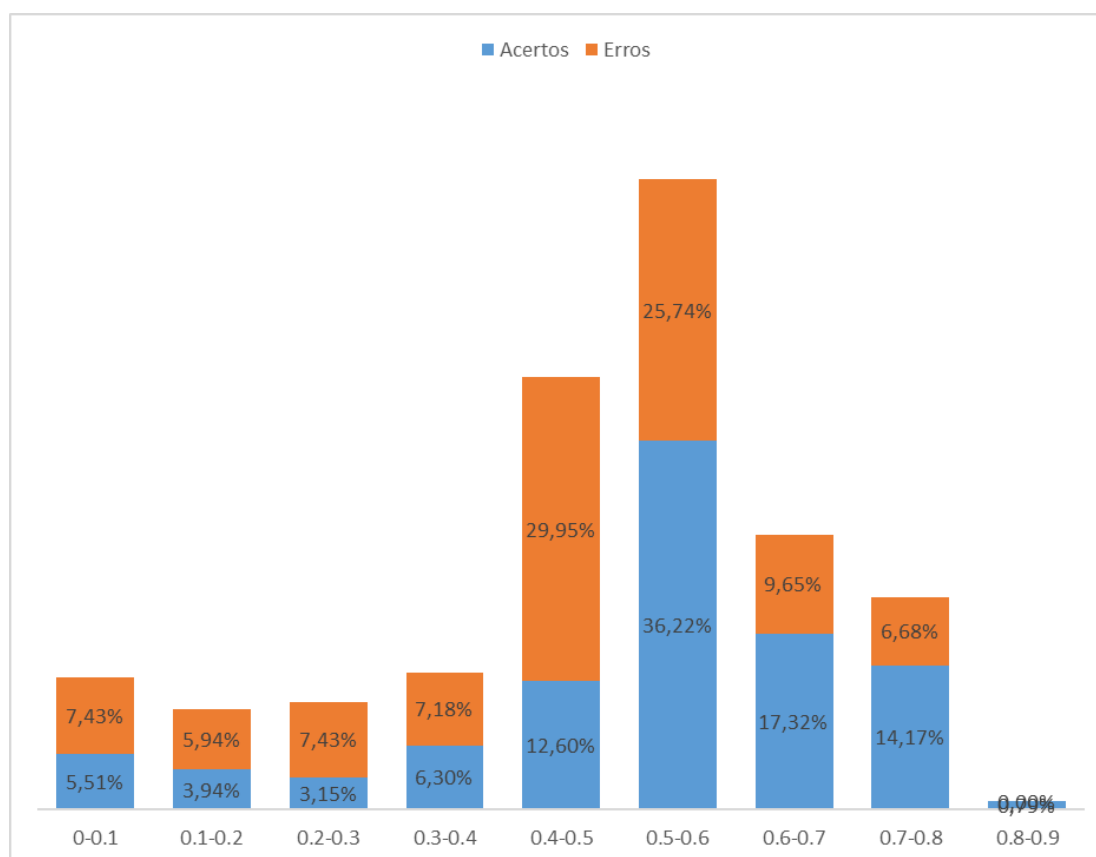
Gráfico 3: Frequência de ocorrências das similaridades entre a resposta do chatbot e a alternativa que tem maior similaridade (disciplina de matemática)



Fonte: O autor.

O Gráfico 4 apresenta a frequência de ocorrências das similaridades entre a resposta do chatbot e a alternativa correta, onde destaca-se com maior frequência o intervalo de 0.5 e 0.6, a média é de 0.459491386.

Gráfico 4: Frequência de ocorrências das similaridades entre a resposta do chatbot e a alternativa correta (disciplina de matemática)



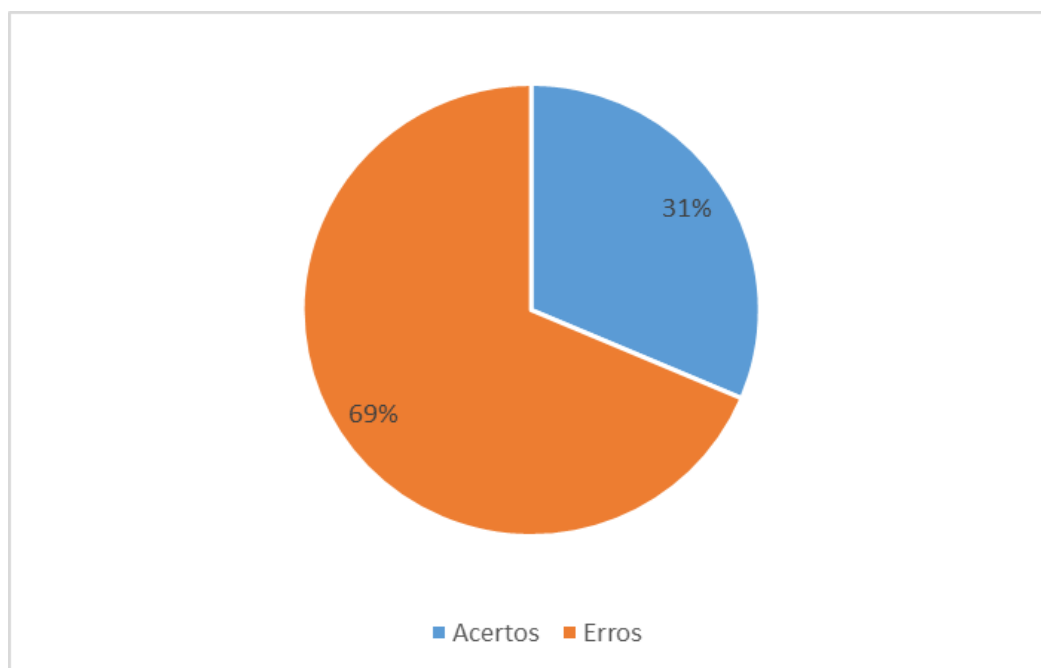
Fonte: O autor.

A eficácia do chatbot na disciplina de matemática foi menos que nas matérias de português e informática, porém maior que medicina, mas como os acertos estão abaixo de 70%, faz com que o chatbot seja ineficaz nessa disciplina.

II. Pedagogia

O dataset de pedagogia contém um total de 2175 questões, onde 892 (80%) questões foram utilizadas para treinamento e 223 (20%) para teste. Pode ser visto no Gráfico 5 que o chatbot obteve um total de 31% de acertos e 69% de erros, sendo assim, estando abaixo somente dos acertos da disciplina de português.

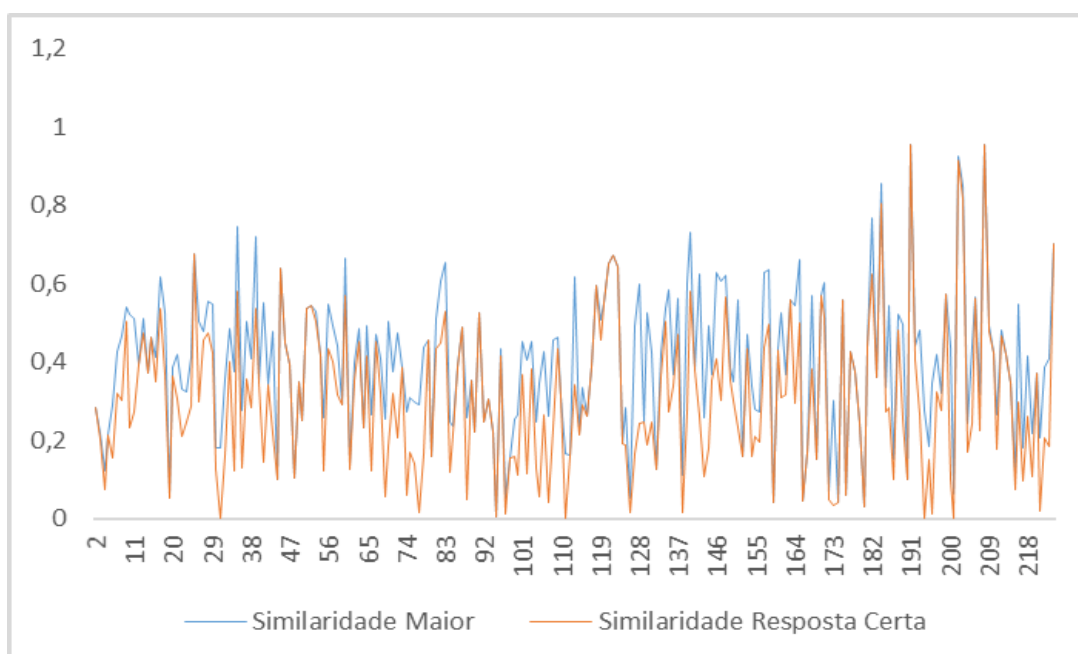
Gráfico 5: Acertos versus erros (disciplina de pedagogia)



Fonte: O autor.

No Gráfico 6, observa-se que as similaridades não chegaram muito perto do valor 1, o que significa que as repostas do chatbot não tiveram muita igualdade com as alternativas.

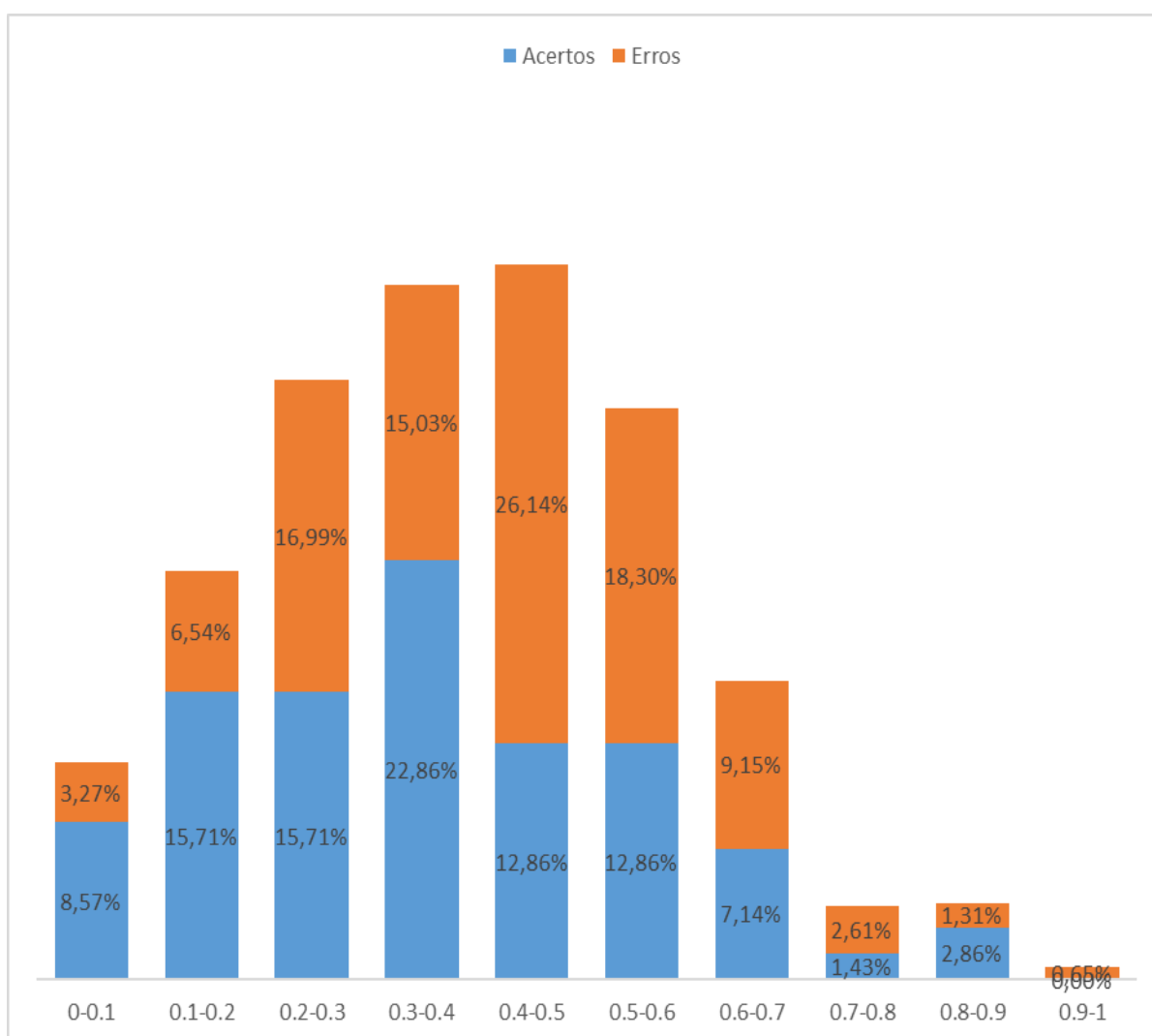
Gráfico 6: Similaridade resposta maior versus similaridade resposta certa (disciplina de pedagogia)



Fonte: O autor.

No Gráfico 7 pode-se observar que a maior frequência de similaridade ocorreu entre valores 0.4 e 0.5, o que significa que ao escolher uma alternativa e destaca-la como correta, na maioria das vezes levou em consideração uma média de similaridade de 0.400259141.

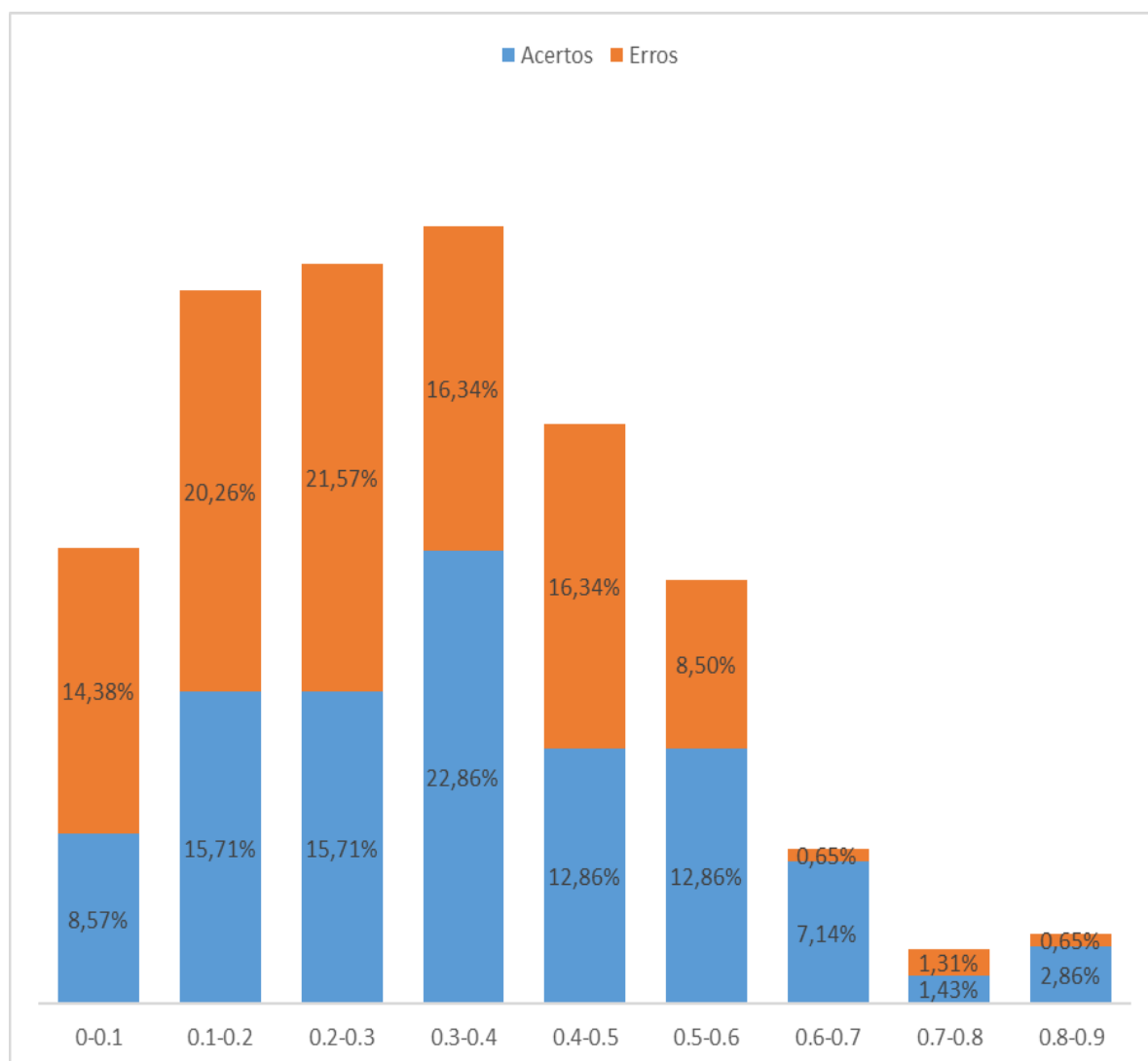
Gráfico 7: Frequência de ocorrências das similaridades entre a resposta do chatbot e a alternativa que tem maior similaridade (disciplina de pedagogia)



Fonte: O autor.

O Gráfico 8 apresenta a frequência de ocorrências das similaridades entre a resposta do chatbot e a alternativa correta, onde destaca-se com maior frequência o intervalo de 0.3 e 0.4, a média é 0.3077488.

Gráfico 8: Frequência de ocorrências das similaridades entre a resposta do chatbot e a alternativa correta (disciplina de pedagogia)



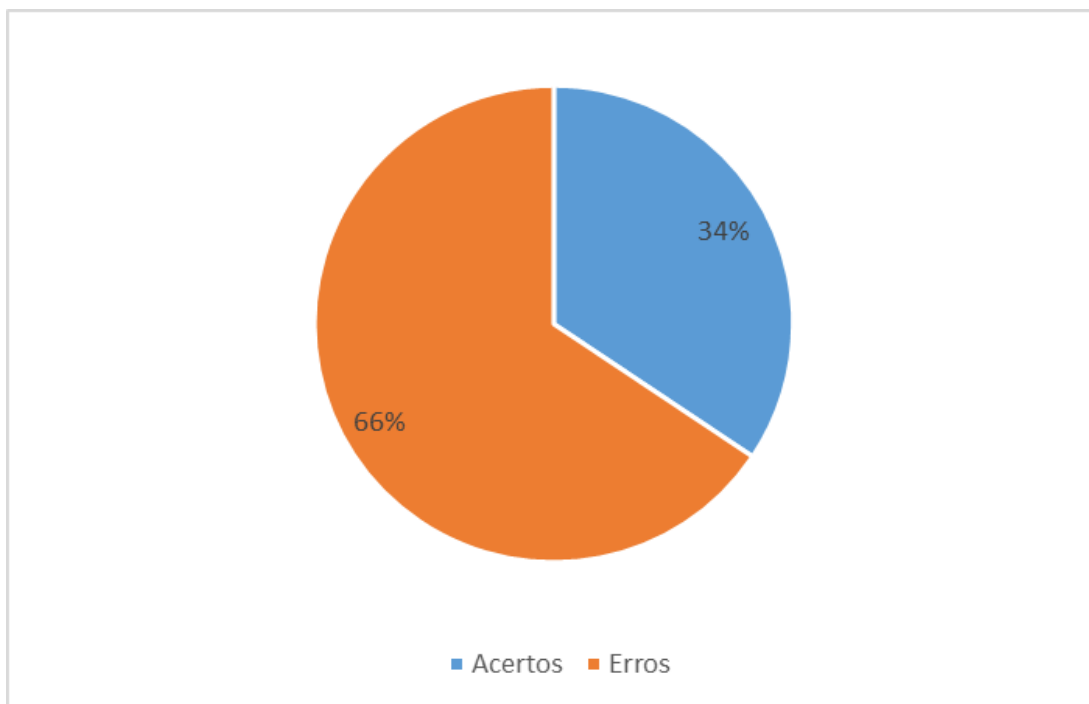
Fonte: O autor.

Pode ser visto que o chatbot se saiu melhor na matéria de pedagogia se comparado as disciplinas: informática, medicina e matemática, perdendo somente para português.

III. Enfermagem

O dataset de português contém um total de 1555 questões, onde 1244 (80%) questões foram utilizadas para treinamento e 680 (20%) para testes. Pode-se observar o Gráfico 9 que o chatbot conseguiu acertar 34% das questões, sendo assim a maior taxa de acertos até o momento, errando 66%.

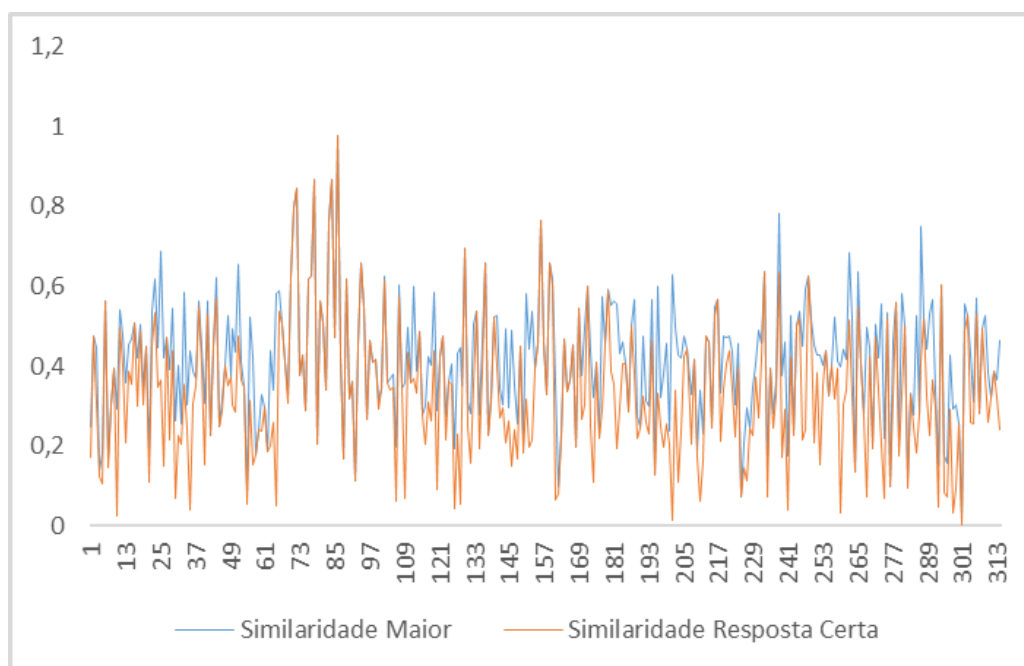
Gráfico 9: Acertos versus erros (disciplina de enfermagem)



Fonte: O autor.

No Gráfico 10 é visto que o chatbot conseguiu responder perguntas tendo 1 de similaridade, ou seja exatamente igual à resposta.

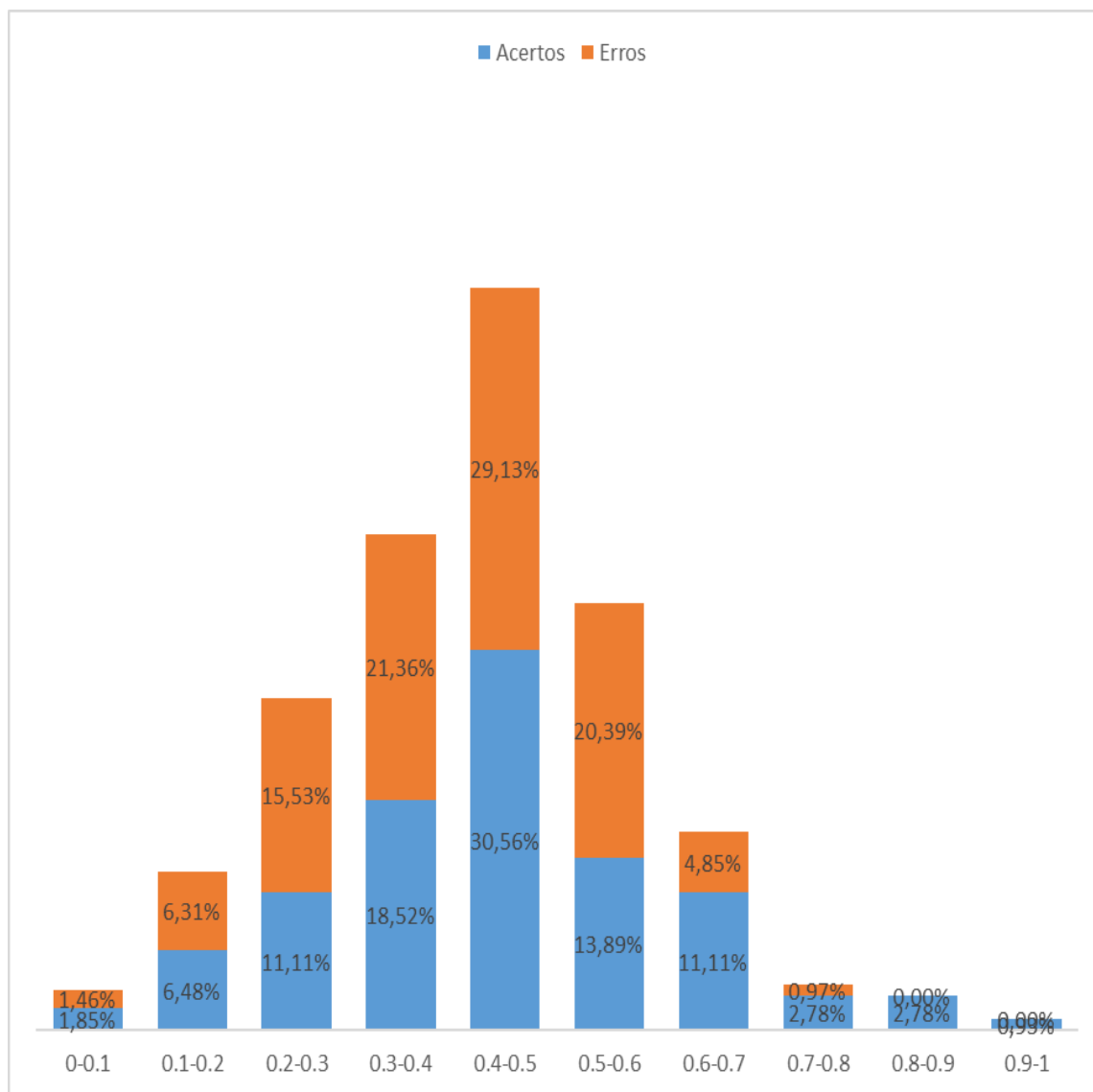
Gráfico 10: Similaridade resposta maior versus similaridade resposta certa (disciplina de enfermagem)



Fonte: O autor.

O Gráfico 11 mostra que a frequência de ocorrências das similaridades entre a resposta do chatbot e a alternativa que tem maior similaridade aconteceu entre 0.4 e 0.5 onde teve 35% de acertos. E com uma média de similaridade de 0.418263305

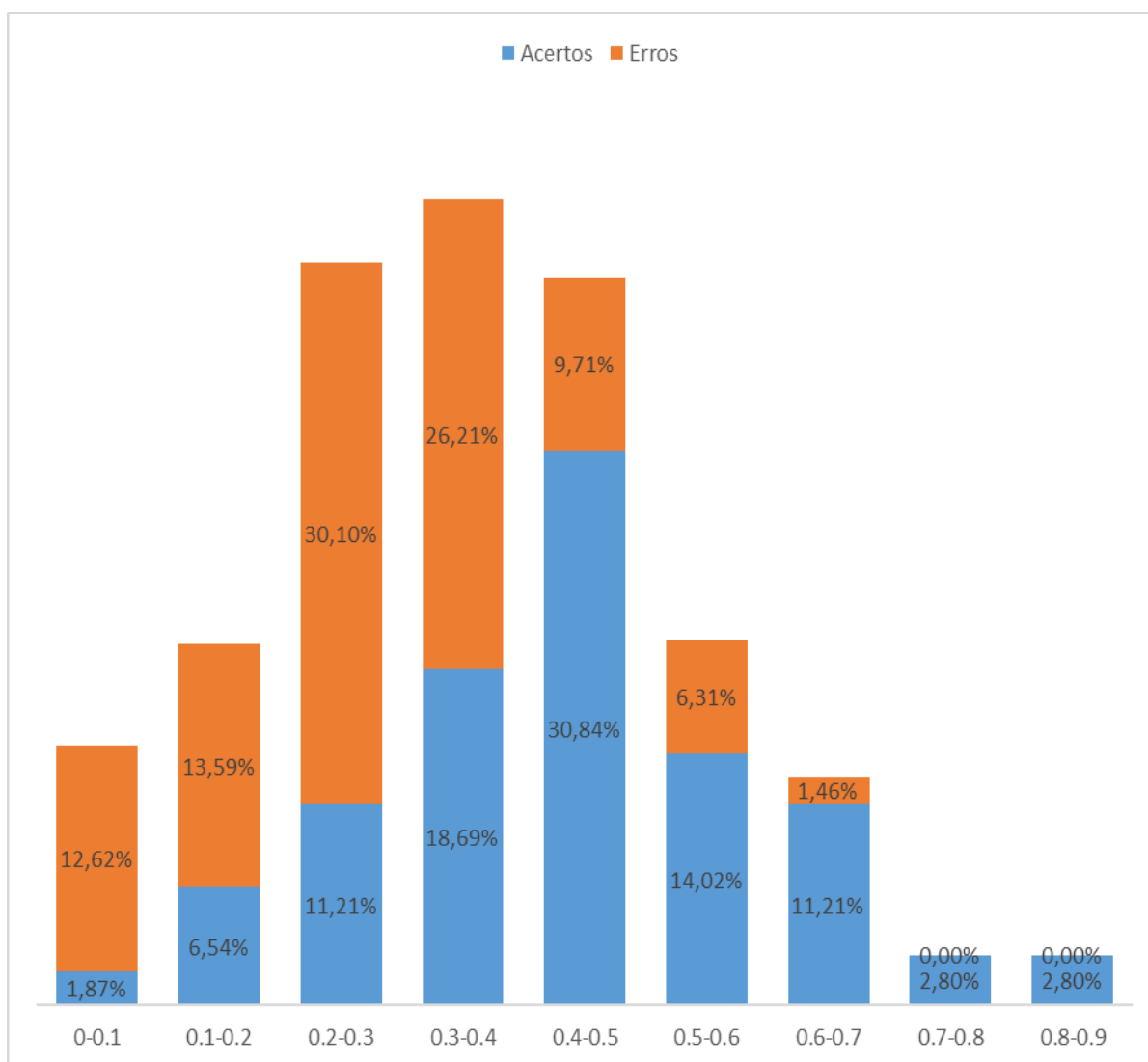
Gráfico 11: Frequência de ocorrências das similaridades entre a resposta do chatbot e a alternativa que tem maior similaridade (disciplina de enfermagem)



Fonte: O autor.

O Gráfico 12 apresenta a frequência de ocorrências das similaridades entre a resposta do chatbot e a alternativa correta, onde destaca-se com maior frequência o intervalo de 0.3 e 0.4 e média em 0.334833596.

Gráfico 12: Frequência de ocorrências das similaridades entre a resposta do chatbot e a alternativa correta (disciplina de enfermagem)

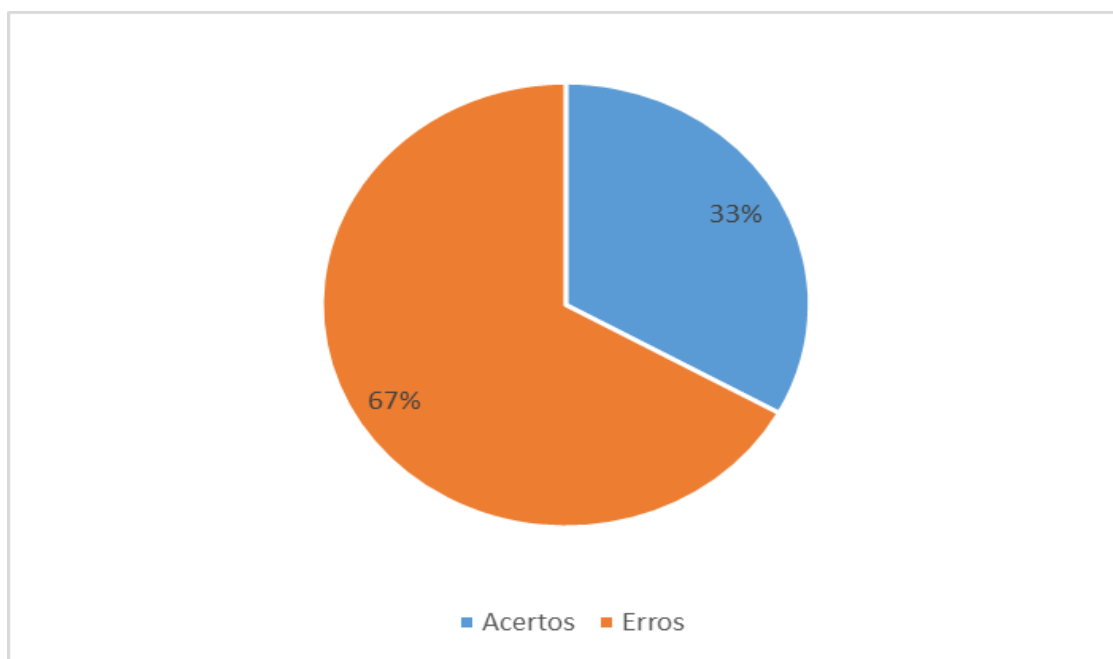


A porcentagem de acertos e a similaridade entre a resposta certa e a resposta do chatbot tiveram aproveitamento de 34%, sendo o maior até o momento, porém não ultrapassa 70% para ser considerado eficaz, e teve média de similaridade baixa.

IV. Inglês

O dataset de inglês contém um total de 3400 questões, onde 2720 (80%) questões foram utilizadas para treinamento e 680 (20%) para teste. Pode ser visto no Gráfico 13, que o chatbot obteve um total de 33% acertos e 67% erros, ficando abaixo somente da disciplina de enfermagem.

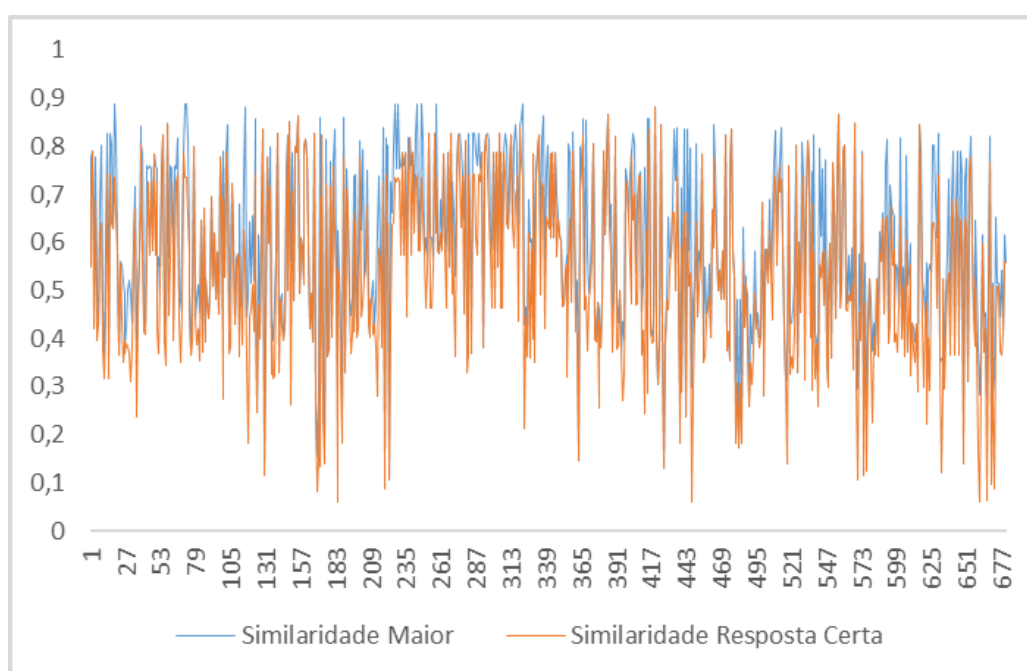
Gráfico 13: Acertos versus erros (disciplina de inglês)



Fonte: O autor.

No Gráfico 14, pode-se observar que a similaridade com a resposta certa muitas vezes se aproximou do valor 0, o que torna as respostas do chatbot diferentes das alternativas corretas

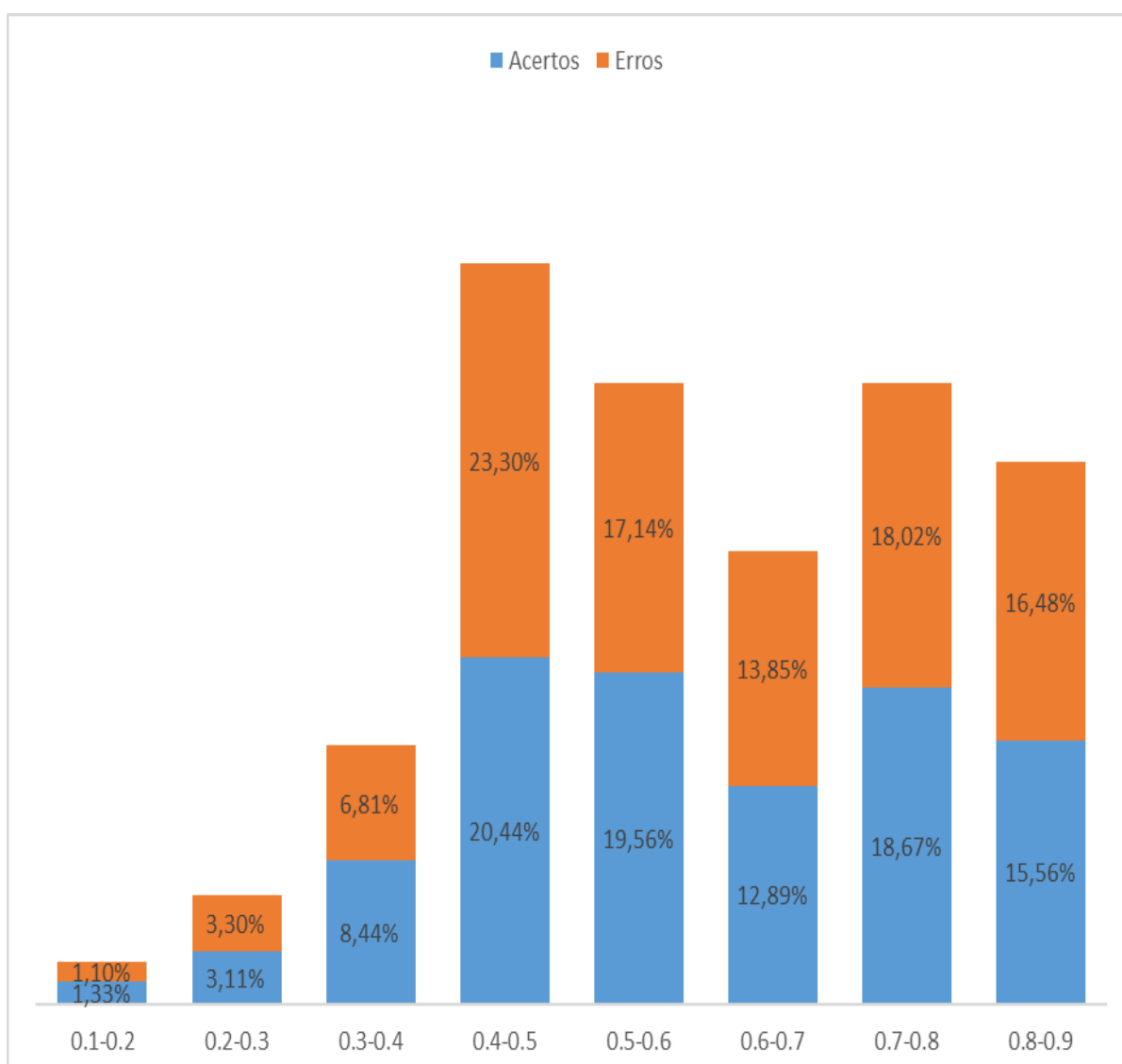
Gráfico 14: Similaridade resposta maior versus similaridade resposta certa (disciplina de inglês)



Fonte: O autor.

A média de similaridade entre a resposta do chatbot e a alternativa que tem maior similaridade é de 0.597460487, porem conforme pode ser visto no Gráfico 15, a maior frequência aconteceu entre 0.4 e 0.5, e também pode ser visto que a disciplina de inglês é a que representa similaridades mais altas dentre todas as disciplinas.

Gráfico 15: Frequência de ocorrências das similaridades entre a resposta do chatbot e a alternativa que tem maior similaridade (disciplina de inglês)

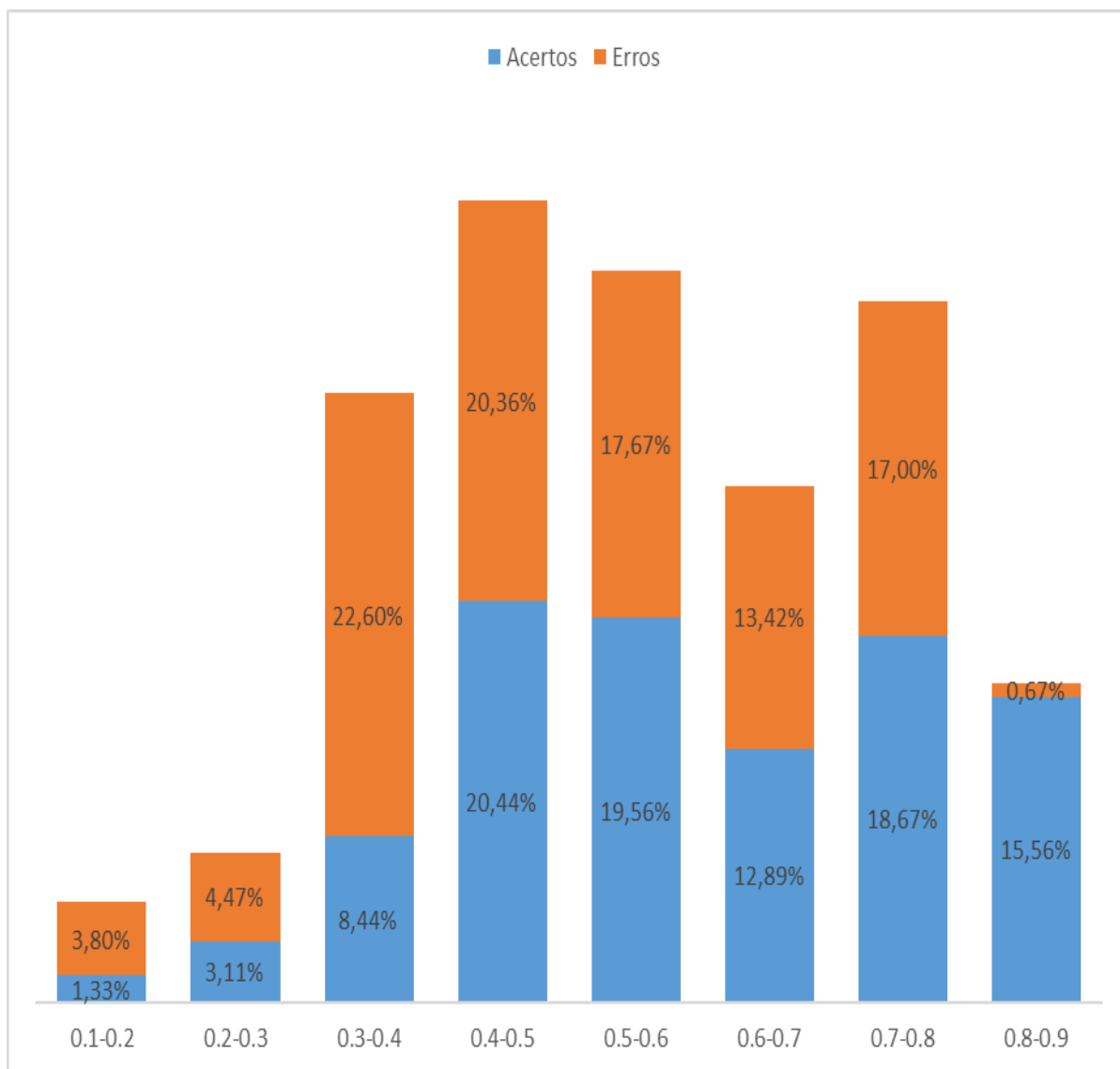


Fonte: O autor.

Já a média de similaridade entre a resposta do chatbot e a alternativa correta é de 0.531113171, e conforme pode ser visto no Gráfico 16 **Erro! Fonte de referência não encontrada.**, a maior frequência também aconteceu entre 0.4 e 0.5. pode-se

notar que no intervalo de 0.8 e 0.9 teve uma boa frequência de similaridade nas respostas certas

Gráfico 16: Frequência de ocorrências das similaridades entre a resposta do chatbot e a alternativa correta (disciplina de inglês)



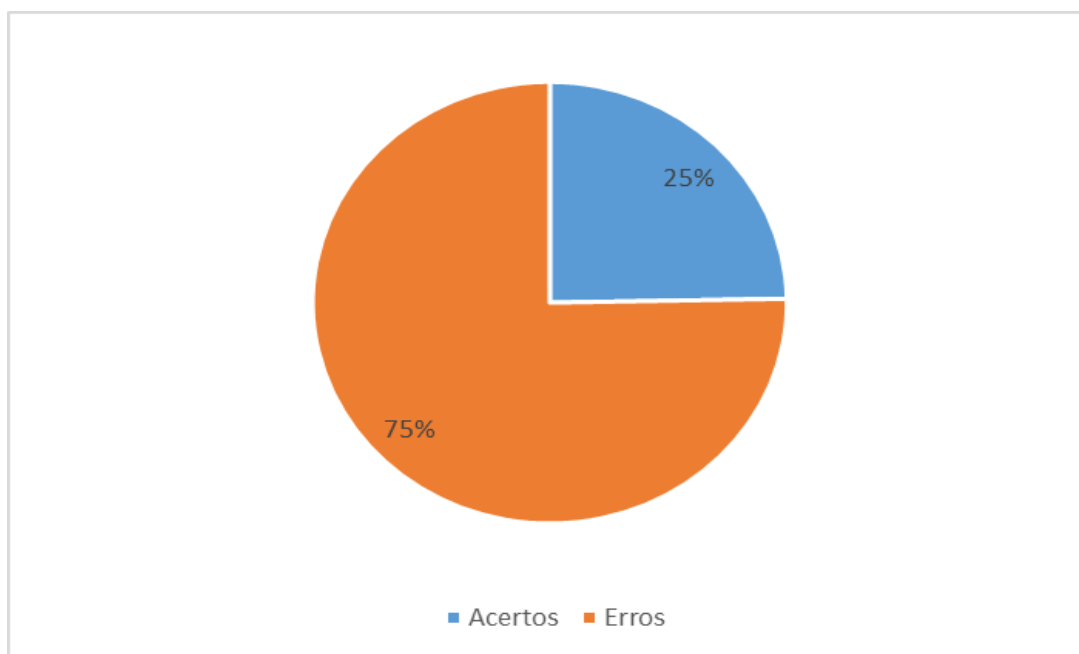
Fonte: O autor.

Na disciplina de inglês o chatbot conseguiu chegar mais próximo das alternativas, porém obteve um aproveitamento de 33%, e estando abaixo da disciplina de enfermagem sendo assim considerado ineficaz.

V. Raciocínio Lógico

O dataset contém um total de 1780 questões, onde 1424 (80%) questões foram utilizadas no treinamento e 356 (20%) foram utilizadas para teste. Podemos observar no Gráfico 17 **Erro! Fonte de referência não encontrada.** que o chatbot obteve um total de 25% de acertos e 75% de erros.

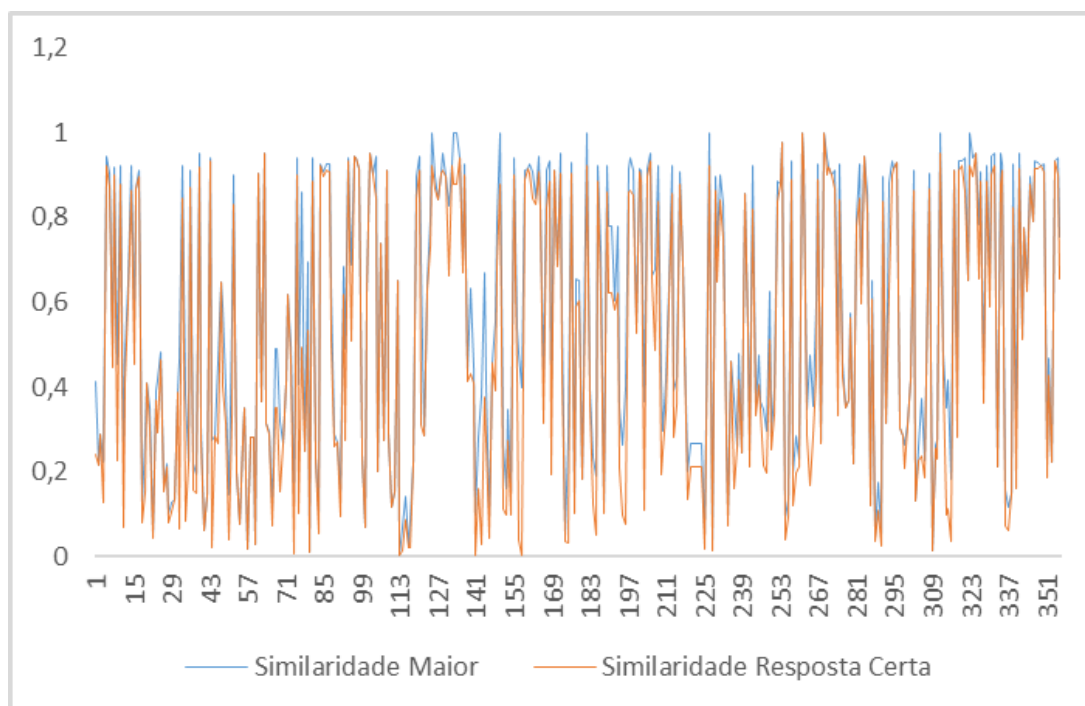
Gráfico 17: Acertos versus erros (disciplina de raciocínio lógico)



Fonte: O autor.

No Gráfico 18Gráfico, é mostrado que na matéria de raciocínio logico é existe a maior oscilação em relação as similaridades.

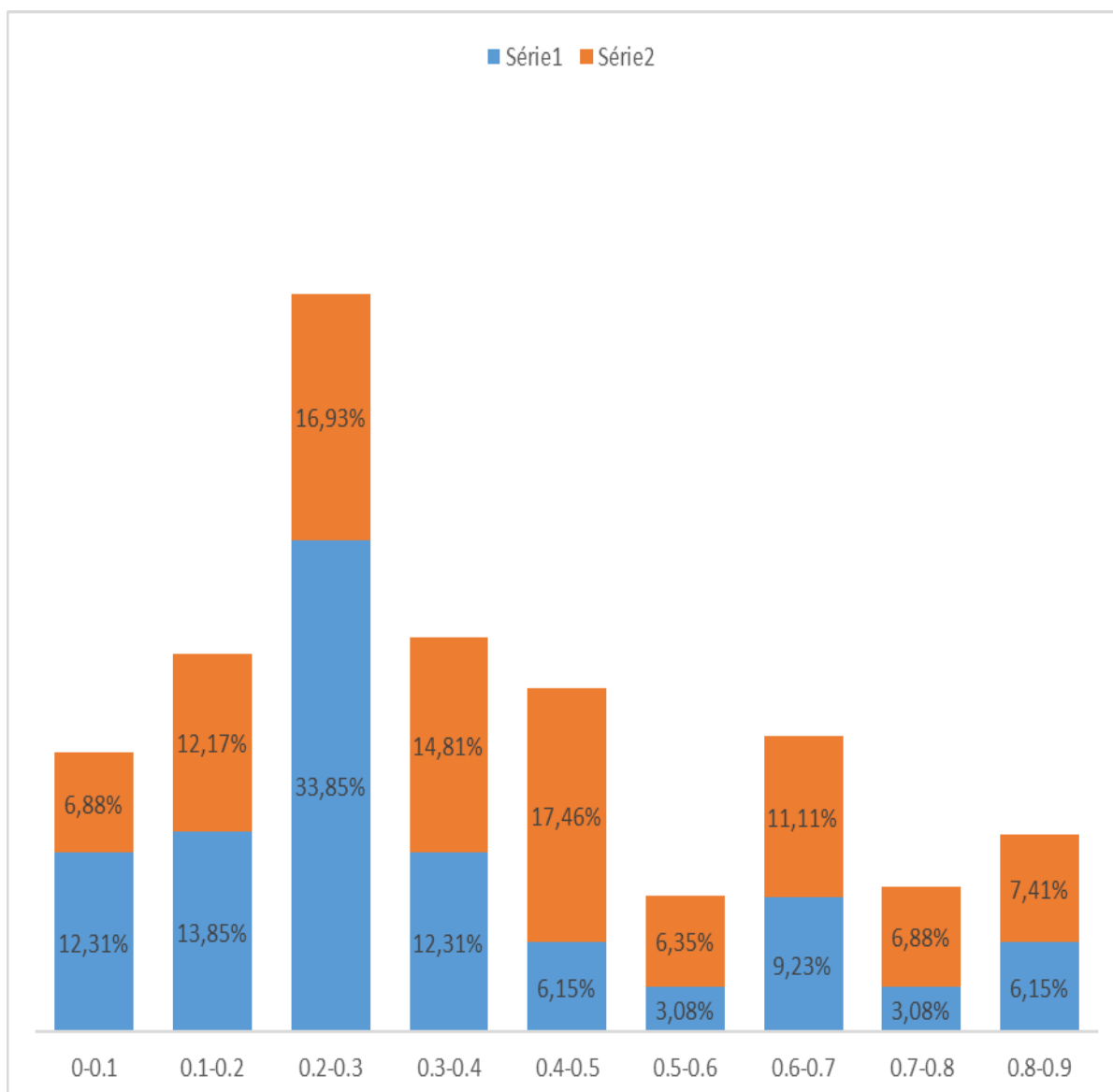
Gráfico 18: Similaridade resposta maior versus similaridade resposta certa (disciplina de raciocínio lógico)



Fonte: O autor.

No Gráfico 19 **Erro! Fonte de referência não encontrada.** pode-se observar que a maior frequência de similaridade ocorreu entre 0,2 e 0,3, o que significa que ao escolher uma alternativa e destaca-la como correta, na maioria das vezes levou em consideração uma média de similaridade de 0,556007098.

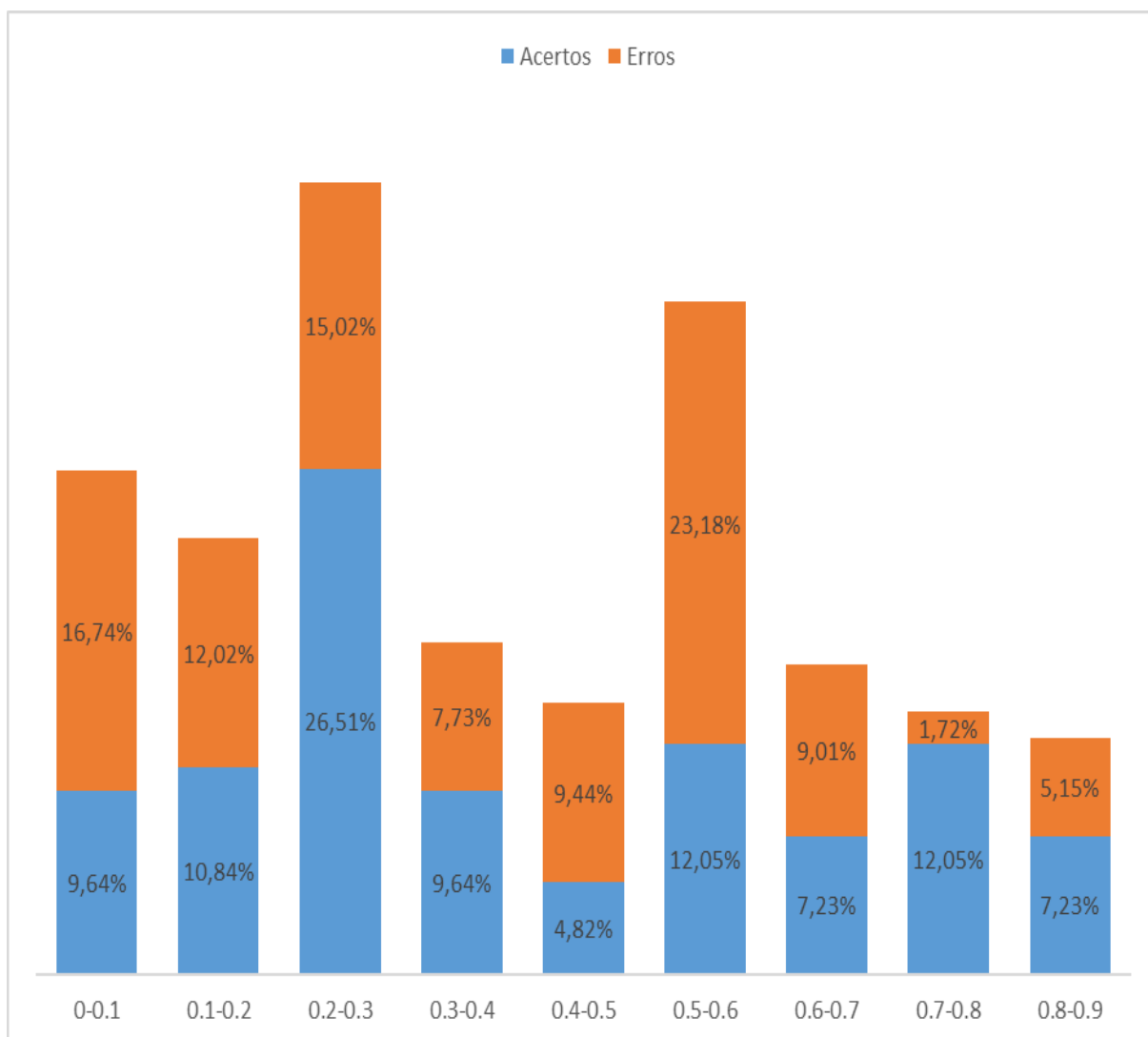
Gráfico 19: Frequência de ocorrências das similaridades entre a resposta do chatbot e a alternativa que tem maior similaridade (disciplina de raciocínio lógico)



Fonte: O autor.

O Gráfico 20 apresenta a frequência de ocorrências das similaridades entre a resposta do chatbot e a alternativa correta, e também tem a maior frequência no intervalo de 0.2 e 0.3, a média é de 0.494397363.

Gráfico 20: Frequência de ocorrências das similaridades entre a resposta do chatbot e a alternativa correta (disciplina de raciocínio lógico)



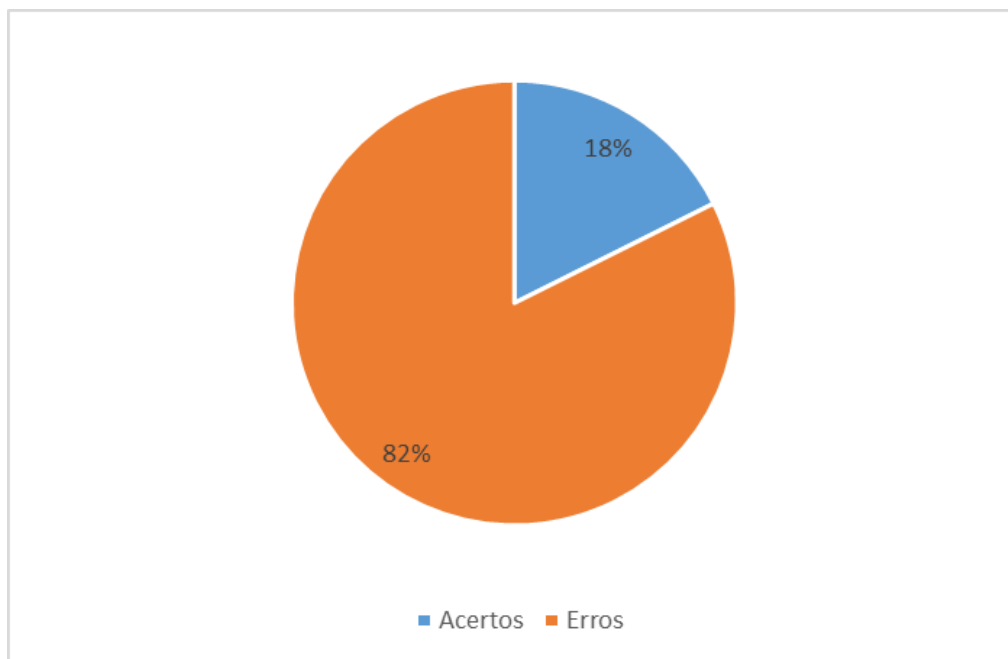
Fonte: O autor.

A porcentagem de acertos e a similaridade entre a resposta certa e a resposta do chatbot foram abaixo de 70%, sendo assim considerado ineficaz.

VI. Economia

O dataset de economia contém um total de 1220 questões, onde 976 (80%) questões foram utilizadas para o treinamento e 224 (20%) para teste. Pode-se observar no Gráfico 21 **Erro! Fonte de referência não encontrada.** que o chatbot obteve um total de 18% de acertos e 82% de erros

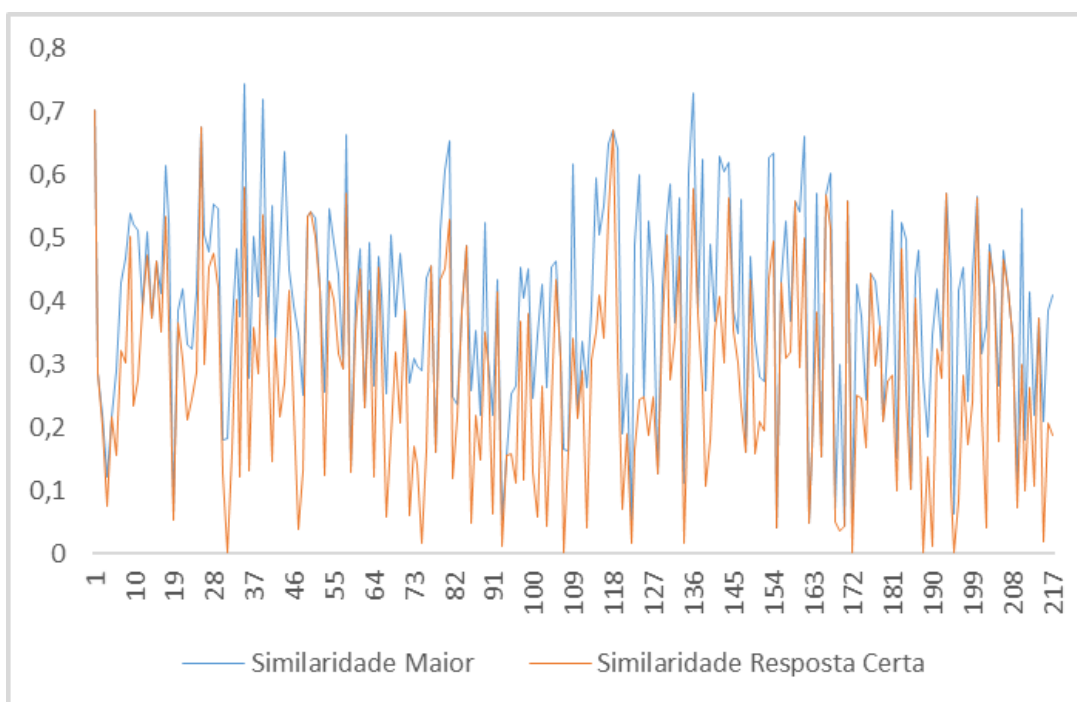
Gráfico 21: Acertos versus erros (disciplina de economia)



Fonte: O autor.

No Gráfico 22 **Erro! Fonte de referência não encontrada.** é mostrado que as similaridades ficaram muito dispersas, se comparado com as disciplinas anteriores.

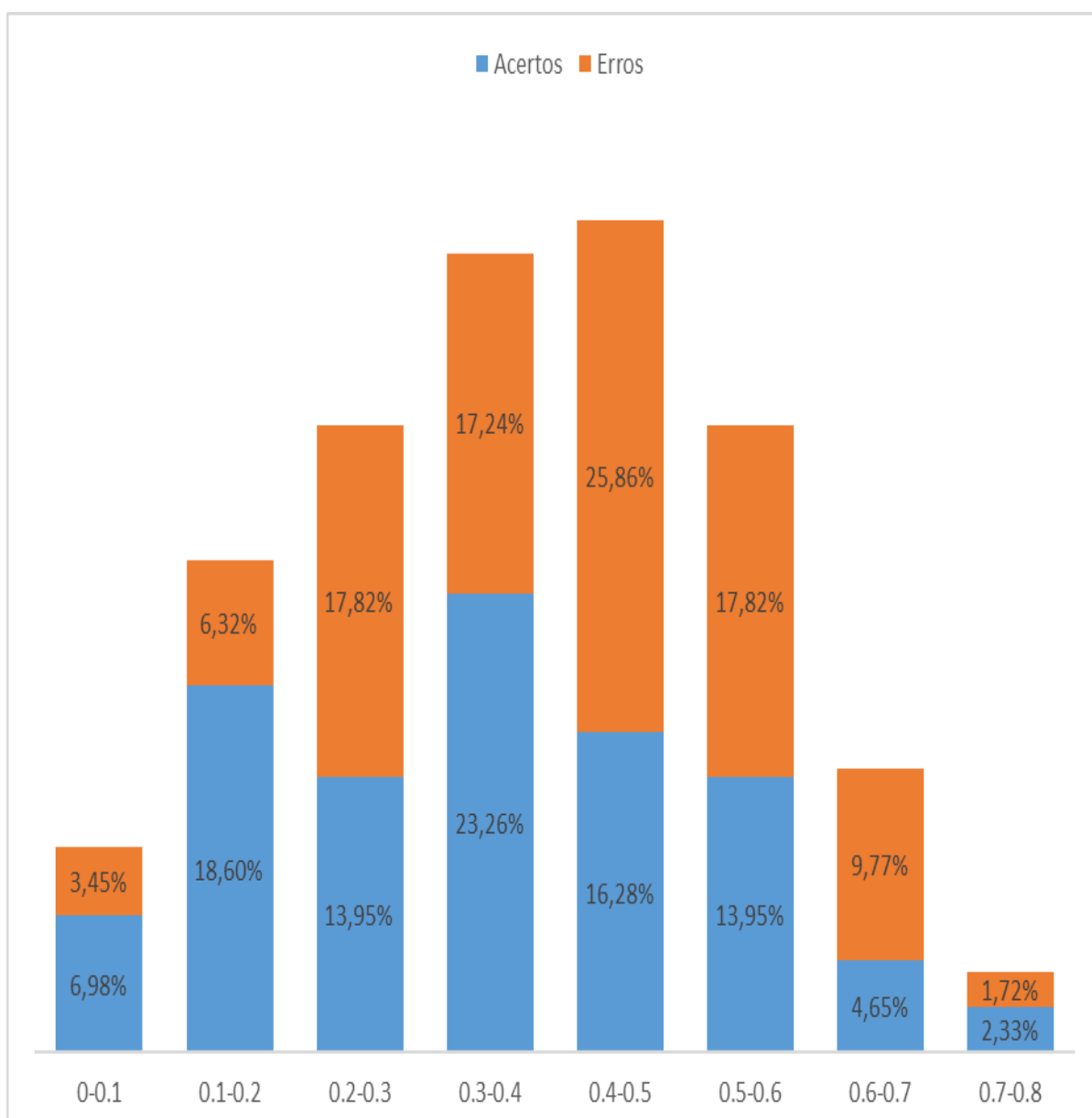
Gráfico 22: Similaridade resposta maior versus similaridade resposta certa (disciplina de economia)



Fonte: O autor.

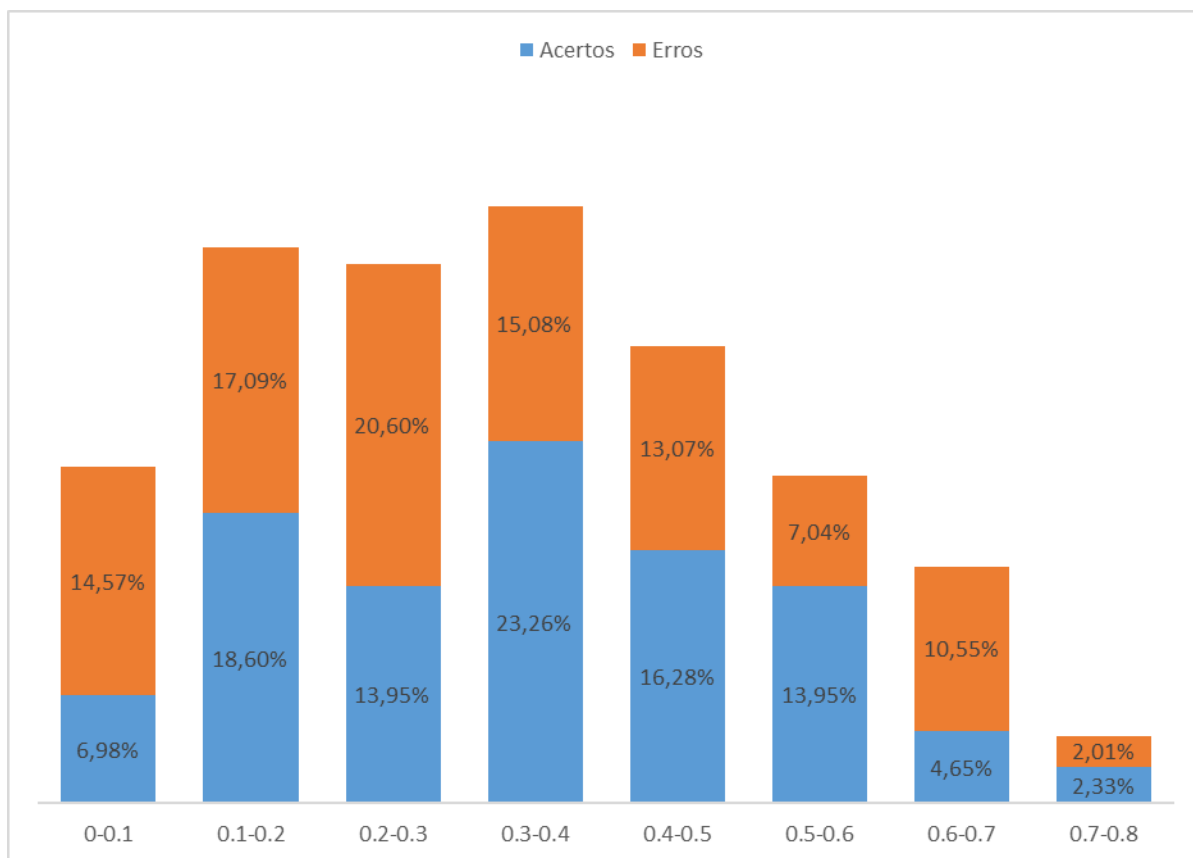
No Gráfico 23 **Erro! Fonte de referência não encontrada.** pode-se observar que a maior frequência de similaridade ocorreu entre valores 0.4 e 0.5, e com média de 0.393294972, o que significa que ao escolher uma alternativa e destaca-la como correta, na maioria das vezes levou em consideração uma média de similaridade de 0.393294972.

Gráfico 23: Frequência de ocorrências das similaridades entre a resposta do chatbot e a alternativa que tem maior similaridade (disciplina de economia)



Já no Gráfico 24, pode ser visto que a maior similaridade foi entre 0.3 e 0.4, ou seja, não teve um aproveitamento maior que 70%, e uma média de 0.27951265.

Gráfico 24: Frequência de ocorrências das similaridades entre a resposta do chatbot e a alternativa correta (disciplina de economia)




Fonte: O autor.

Na disciplina de economia a taxa de acertos não foi superior a 70%, sendo assim considerada baixa.

APÊNDICE B

AUTORIZAÇÃO DE DIVULGAÇÃO DE DADOS

Pelo presente, em nome da VERSATILE CONSULTORIA E TECNOLOGIA DA INFORMACAO inscrita sob o CNPJ nº 17.200.610/0001-31, a qual represento neste ato enquanto diretor de tecnologia, autorizo Cleiton Messias da Silva Lopes a iniciar um estudo de caso para fins acadêmicos para a Rede de Ensino Doctum, utilizando nossa base de dados de questões acadêmicas de diversas áreas do conhecimento humano. O uso e divulgação desses dados limita-se a exposição em seu Trabalho de Conclusão de Curso, além de ser aplicável apenas ao tempo necessário para que este seja finalizado, não permitindo sua divulgação, cópia, distribuição ou acesso a terceiros por quaisquer meios eletrônicos ou não.



Glauber Costa

Diretor de tecnologia

VersaTec