Ministério da Educação
Universidade Tecnológica Federal do Paraná
Campus Londrina

APLICAÇÃO DE TÉCNICAS DE APRENDIZADO DE MÁQUINA PARA CLASSIFICAR ALUNOS DE CURSOS DE IDIOMAS COM RELAÇÃO À POSSIBILIDADE DE EVASÃO

Londrina
2019

MONIQUE TAMARA DE LIMA

APLICAÇÃO DE TÉCNICAS DE APRENDIZADO DE MÁQUINA PARA CLASSIFICAR ALUNOS DE CURSOS DE IDIOMAS COM RELAÇÃO À POSSIBILIDADE DE EVASÃO

Trabalho de Conclusão de Curso de graduação, apresentado à disciplina TCC 2, do curso de Engenharia de Produção da Universidade Tecnológica Federal do Paraná – UTFPR- câmpus Londrina, como requisito parcial para a obtenção do título de Bacharel.

Orientador: Dr. Rafael Henrique Palma Lima.

Londrina

201

TERMO DE APROVAÇÃO

TÍTULO DA MONOGRAFIA

POR

MONIQUE TAMARA DE LIMA

Esta Monografia foi apresentada às 16 horas do dia 27 de junho de 2019 como requisito parcial para obtenção do título de bacharel em ENGENHARIA DE PRODUÇÃO, Universidade Tecnológica Federal do Paraná – Câmpus Londrina. O candidato foi arguido pela Banca Examinadora composta pelos professores relacionados abaixo. Após deliberação, a Banca Examinadora considerou o trabalho: **APROVADO**.

Prof. Dr. Bruno Samways Dos Santos
Banca Examidadora

Prof. Dr. Rogerio Tondato
Banca Examidadora

Prof. Dr. Rafael Henrique Palma Lima

Presidente da Banca Examinadora
Orientador

RESUMO

O aumento da competitividade no mercado de trabalho tem levado as pessoas a buscarem novas habilidades e conhecimentos, dentre os quais se destacam os cursos de idiomas. Entretanto os cursos vêm sofrendo com os altos índices de evasão, sendo este causado por múltiplos fatores, os quais possuem relacionamentos complexos que dificultam a elaboração de modelos de classificação. Sob este contexto, a atual pesquisa propõe a utilização de técnicas de Aprendizado de Máquina capazes de analisar grandes e complexos bancos de dados, tendo como propósito identificar antecipadamente um aluno que seja propenso a evadir do curso de idioma, possibilitando assim a tomada de medidas para reduzir a taxa de evasão. Para estudar esse problema foi realizado um levantamento na literatura sobre os fatores que podem causar a evasão dos alunos e, em seguida, elaborou-se um questionário usando o *Google Forms*. Os questionários foram pré-processados e 7 técnicas de aprendizado de máquina foram utilizadas no estudo de dois modelos de classificação, cada um com duas configurações distintas. O primeiro modelo visava buscava prever se o aluno estava cursando, se havia evadido ou se tinha concluído o curso de idioma, enquanto o segundo tinha como propósito prever apenas se o estudante evadiu ou não. Os resultados foram satisfatórios, com destaque para as técnicas Máquina de Vetores de Suporte e Florestas Aleatórias que obtiveram um índice máximo de acuracidade de 91% e 88%, respectivamente.

Palavras-chaves: Aprendizado de Máquina. Inteligência Artificial. Evasão de Cursos de Idiomas. Técnicas de Classificação.

ABSTRACT

The increase in competitiveness in the labor market has led people to seek new skills and knowledge, among which are the language courses. However, the courses have been suffering from high dropout rates, which is caused by multiple factors, which have complex relationships that make difficult the elaboration of classification models. In this context, the current research proposes the use of Machine Learning techniques capable of analyzing large and complex databases, with the purpose of identifying in advance a student who is prone to evade the language course, thus enabling measures to be taken reduce the rate of evasion. In order to study this problem, a literature review was carried out on the factors that can cause student drop-outs, and then a questionnaire was developed using Google Forms. The questionnaires were pre-processed and 7 machine learning techniques were used in the study of two classification models, each with two different configurations. The first model aimed to predict whether the student was attending, whether he had escaped or had completed the language course, while the second was intended to predict only whether the student evaded or not. The results were satisfactory, with emphasis on the techniques of Support Vector Machines and Random Forests, which obtained a maximum accuracy of 91% and 88%, respectively.

Key-words: Machine Learning. Artificial intelligence. Evasion of Language Courses. Classification Techniques.

LISTA DE ILUSTRAÇÕES

Figura 1 - Viés e variação no lançamento de dardos 23
Figura 2 - Ilustração de Sobre-ajuste, Sub-ajuste 24
Figura 3 - Hierarquia do aprendizado de máquina supervisionado e não supervisionado 25
Figura 4 - Processo de aplicação de um problema de Aprendizado de Máquina.....26
Figura 5 - Exemplo de um problema de projetar pontos 2D em uma dimensão 29
Figura 6 - Matriz de confusão separada em quatro quadrantes: Verdadeiro positivo; Falso positivo; Falso negativo; Verdadeiro negativo 30
Figura 7 - Exemplo de classificação por meio do KNN 31
Figura 8 - Representação da função sigmoide.....32
Figura 9 - Exemplo de um problema bidimensional separável, na qual as instâncias mais próximas do hiperplano de margem máxima, denominadas de vetores de suporte foram marcados com quadrados quadriculados.....34
Figura 10 - Exemplo de aplicação do método árvore de decisão para a classificação de regiões.....35
Figura 11 - Exemplo de classificação utilizando o modelo de Florestas Aleatórias...37
Figura 12 - Exemplo de uma Rede de Multicamadas.....39
Figura 13 - Fluxograma das etapas da pesquisa 43
Figura 14- Taxa de erro versus o K número de vizinhos.....48
Figura 15 - Scree plot dos porcentuais de variância explicada 50
Figura 16 – *Biplot* de todos os componentes 51
Figura 17 – Percentual de entrevistado por “Caso”.....53
Figura 18 - Percentual de entrevistado por modalidade de estudo 54
Figura 19 – Perfil dos respondentes de acordo com o sexo e a renda familiar.....55
Figura 20 - Histograma com a idade atual 55
Figura 21 Histograma com idade do participando de quando foi realizado o curso. 56
Figura 22 - Boxplot.....56
Figura 23 - Mapa de calor subdivido de acordo com os clusters.....57
Figura 24 – Percentuais de acuracidade das técnicas referente ao modelo 1A.....62
Figura 25 - Percentuais de acuracidade das técnicas referente ao modelo 1B.....63
Figura 26 - Percentuais de acuracidade das técnicas referente ao modelo 2A.....64

Figura 27 - Percentuais de acuracidade das técnicas referente ao modelo 2B.....65
Figura 28 - Comparativo dos percentuais de acuracidade obtidos pelas técnicas de aprendizado de máquina.....66
Figura 29 - Matriz de confusão dos valores reais versus os preditos na fase de teste 67

LISTA DE TABELAS

Tabela 1 - Síntese dos aspectos que contribuem para uma possível evasão escolar e seus respectivos pesquisadores 20
Tabela 2 - Um exemplo da codificação das variáveis de raça, codificadas em três níveis 33
Tabela 3 - Percentual de nível de dificuldade experimentada pelos estudantes de acordo com a classe. 58
Tabela 4 - Índices percentuais relativos as faixas etárias de idades que os estudantes possuíam quando realizaram o curso. 59
Tabela 5 - Percentual de respondentes referente ao tempo de estudo..... 59
Tabela 6 - Percentual quanto ao nível de escolaridade dos genitores dos respondentes..... 59
Tabela 7 - Percentual entre o Idioma estudo e classe dos respondentes. 60
Tabela 8 - Situações que ocorreram durante a realização do curso de idioma..... 60
Tabela 9 – Dificuldades com locomoção, matérias didáticos e comunicação com o tutor. 61
Tabela 10 - Resultados referentes aos Modelo 2A e 2B 67
Tabela 11 - Sugestões de medidas que podem ser tomadas pelas instituições de ensino de idiomas 69

SUMÁRIO

1.INTRODUÇÃO 10

1.1. Objetivo Geral 12

1.2. Estrutura do trabalho 12

2. REFERENCIAL TEÓRICO 14

2.1. Levantamento dos principais aspectos causadores da evasão escolar 14

2.2. Aprendizado de máquina 21

2.2.1. Aprendizado de máquina supervisionado 26

2.2.2. Aprendizado de máquina não supervisionado 27

2.3. Aplicação do aprendizado de máquina na evasão escolar 27

2.4. Técnicas de aprendizado de máquina 28

2.4.1. Análise de componentes principais 28

2.4.2. K-vizinhos mais próximos 30

2.4.3. Regressão logística 32

2.4.4. Máquina de vetores de suporte 33

2.4.5. Árvore de decisão 35

2.4.6. Florestas aleatórias 36

2.4.7. Redes neurais 38

3. MÉTODO DE PESQUISA 41

3.1. Etapas da Pesquisa 41

3.2. Coleta de dados e estrutura do questionário 43

3.3. Descrição dos modelos de análise 45

3.4. Implementação de técnicas de aprendizado de máquina 46

3.4.1. Implementação da técnica K-vizinhos mais próximos 47

3.4.2. Implementação da técnica de regressão logística 48

3.4.3. Implementação da técnica de máquina de vetores de suporte 48

3.4.4. Implementação da técnica árvore de decisão 49

3.4.5. Implementação da técnica florestas aleatórias 49

3.4.6. Implementação da técnica Análise de componentes principais 49

3.4.7. Implementação da técnica Redes neurais 52

4. RESULTADOS E DISCUSSÕES 53

4.1. Caracterização dos participantes da pesquisa 53

4.2. Comparação e análise dos classificadores 61
4.2.1. Resultados do Modelo 1A 62
4.2.2. Resultados do Modelo 1B 63
4.2.3. Resultados do Modelo 2A 64
4.2.4. Resultados do Modelo 2B 65
4.2.5. Comparativos dos resultados 66
4.2.6. Resultados das técnicas de aprendizado não supervisionado 67
5. IMPLICAÇÕES PRÁTICAS 69
6. CONSIDERAÇÕES FINAIS 73
REFERÊNCIAS 74
APÊNDICE A - Questionário Caso 1 e Caso 2 81
APÊNDICE B - Questionário Caso 3 85
APÊNDICE C - Questionário Caso 4 89
APÊNDICE D - Questionário Caso 5 93

1. INTRODUÇÃO

Com o aumento da competitividade global se faz necessário que os recursos sejam aplicados com eficiência, sendo assim é essencial a capacidade de um negócio compreender as necessidades dos seus clientes. No entanto, segundo Corrêa *et al.* (2016) isso é uma missão muito complexa, pois as necessidades podem ser atribuídas tanto às transformações ocorridas em uma escala mundial, como também de acordo experiências vivida por cada um, ou ainda podem ser influenciada pelas emoções e sentimentos momentâneos.

Já no âmbito acadêmico, é preciso compreender os diferentes interesses dos estudantes, que podem ser de apenas conquistar um certificado, como também o aluno pode ter o real interesse em adquirir as habilidades que honre tal conclusão, sendo assim a instituição deve assegurar que essa expectativa seja atendida. (SOUSA, 2008).

Frente a esse cenário de globalização e com o advento da internet, propicia-se um ambiente favorável para a comunicação e a troca de informações em escala mundial, ocasionando significativas mudanças entre as formas de interação, na qual o domínio de um segundo idioma tornou-se imprescindível para a formação de profissional. (TONDELLI, 2005).

Sidoski (2014) observou em sua pesquisa o crescente aumento da demanda pelo aprendizado de um ou mais idiomas voltados para o mercado de trabalho, destacando-se o inglês por ser a língua mais falada e por ser visto como meio de garantir empregabilidade, já que é usado corriqueiramente no mundo dos negócios.

Surgiu uma série de instituições de ensino de idiomas no país, oferecendo diferentes metodologias e cada vez mais personalizadas, ocasionando uma concorrência altamente acirrada. (SOUZA, 2008), onde o foco das organizações era voltado apenas atrair novos clientes. No entanto, já é notória a importância de retê-los, uma vez que atrair um novo consumidor pode custar até cinco vezes mais do que agradar e manter aquele já existente. (KOTLER; KELLER, 2012).

Contudo, de acordo como o então chefe do Departamento de Secretaria Estadual de Educação do Estado do Paraná, Cassiano Ogliari, a evasão das turmas do Centro de Línguas Estrangeiras Modernas (CELEM) representa aproximadamente uma taxa anual de 50%. (BUDOLA, 2017). Todavia, acredita-se que as altas taxas de

evasões também se estendam para demais instituições de ensino de idiomas, tornando a evasão uma grave problemática nesse âmbito. Por esta razão, algumas instituições estão criando bancos de dados com os históricos dos seus alunos, ciente da necessidade de conhecer melhor o perfil dos seus estudantes.

Mostrando-se notória a importância de transformar dados em informações valiosas e estratégicas em tempo hábil, como sendo um meio de gerar abertura de novas técnicas de apoio à decisão que permitem uma adaptabilidade e capacidade de resposta ágil. (BAZZOTTI; GARCIA, 2006)

Visto a facilidade de acesso à informação, disponibilidade de dados históricos, e o avanço do poder de processamento e seu barateamento, já existem grandes empresas mundiais que utilizam bancos de dados como forma de colher informações de seus clientes a fim de identificar e satisfazer suas necessidades antes da concorrência. Nesse contexto, as empresas que utilizam o aprendizado de máquina possuem uma visão ampla das necessidades demandadas, sendo então as pioneiras em ideias inovadoras, conquistando assim a confiança do mercado. Deste modo, são bons motivos nos levam a crer que a análise Inteligente dos dados se tornará mais comum no progresso tecnológico. (SMOLA, 2008).

Em reportagem postada pelo site *Data Science Academy* (2018), afirma-se que o aprendizado de máquina é uma das tendências mais tecnológicas e modernas da atualidade. Ainda neste editorial, foi citado que a empresa Gartner presumiu que até o ano de 2020, as tecnologias de Inteligência Artificial, inclusive a aprendizagem de máquina, “estarão presentes em quase todos os novos produtos e serviços de *software*”.

Sob esta concepção, o Aprendizado de Máquina vem sendo desenvolvido para tratar de maneira interativa com as oscilações do mercado, trazendo modelos capazes de analisar grandes e complexos bancos de dados, respondendo de maneira ágil, automática, precisa e em um formato compreensível aos gestores afins de apoiar a tomada de decisão, ou ainda, futuramente as decisões poderão ser tomadas sem nenhuma intervenção humana.

Ciente que as elevadas taxas de evasão causam grandes prejuízos não somente aos alunos e as instituições envolvidas, mas também da nação como um todo, visto que a educação é elemento fundamental para o progresso social e econômico .(NERI, 2009). Logo, é justificável a utilização do aprendizado de máquina

para a previsão de evasão, para predizer o comportamento dos seus estudantes, a fim de melhorar a política de *marketing* das instituições, remodelar as diretrizes de ensino, e aperfeiçoar as experiências de aprendizado dos alunos, evitando o desperdício de tempo e recursos.

1.1. Objetivo Geral

O objetivo deste trabalho é desenvolver modelos de classificação da evasão escolar no contexto do ensino de idiomas e testar a eficiência desses modelos utilizando dados coletados a partir de indivíduos que estudam ou estudaram idiomas. Tais modelos serão desenvolvidos utilizando técnicas consolidadas de aprendizado de máquina e avaliadas usando índices de acuracidade. Este objetivo pode ser desdobrado nos seguintes objetivos específicos:

- Indicar a metodologia que melhor se adapta para o estudo em questão;
- Implementar as técnicas uma linguagem computacional capaz de predizer a evasão escolar em um tempo hábil, podendo ser utilizado futuramente na prática.
- Comparar a eficiência do modelo aplicado a partir dos resultados obtidos após a implementação da metodologia e efetuar comparações quanto a acuracidade das técnicas com o intuito de definir o mais indicado para a predição;
- Discutir as contribuições acadêmicas, dado que a abordagem sobre evasão de cursos de idiomas possui uma baixa representatividade de pesquisas científicas. Assim como validar os resultados obtidos pelas técnicas por meio de métricas pré-estabelecidas.

1.2. Estrutura do trabalho

Quanto às etapas metodológicas, as mesmas foram divididas em quatro sessões. Na primeira seção, foi levantado o referencial teórico, que apresenta uma abordagem dos possíveis atributos causadores da evasão escolar, assim como as fundamentações teóricas das técnicas de aprendizado de máquina que foram utilizadas para análises dos dados. Enquanto a segunda sessão descreveu a estrutura do instrumento de coleta de dados, o período e a metodologia para a aplicação dos

questionários, assim como foi descrito os meios utilizados para implementação das técnicas de aprendizado de máquina.

Na terceira sessão foi discutido a caracterização dos participantes, e analisado e comparado o nível de acuracidade dos modelos e do. Já a quarta sessão foi dedicada para a discussão das implicações práticas.

E por fim, a quinta seção teve como propósito apresentar as contribuições sobre o conteúdo estudado e as conclusões obtidas pela a atual pesquisa, bem como foram dadas sugestões para pesquisas futuras sobre o tema.

2. REFERENCIAL TEÓRICO

No presente capítulo será abordado os principais conceitos necessários para o entendimento do trabalho desenvolvido, como aspectos causadores da evasão escolar, aprendizado de máquina e trabalhos correlatos.

2.1. Levantamento dos principais aspectos causadores da evasão escolar

Nesse tópico será discutido a relevância das perceptivas abordadas nos formulários aplicados para a coleta de dados. Deve se frisar que a maioria dos estudos levantados para esse referencial diz respeito à evasão de alunos no ensino superior, contudo é notório que a causas desse evento é similar em qualquer instituição de ensino.

Segundo Oliveira Júnior (2015), foi realizado pelo Ministério da Educação (MEC) um trabalho que tinha como objetivo sistematizar o problema da evasão escolar da nação, sendo criada a Comissão Especial para o Estudo da Evasão nas Universidades Brasileiras. Os objetivos finais dessa Comissão foram elucidar o conceito de evasão, analisar as taxas e as razões desse evento e padronizar uma metodologia a ser utilizadas pelas instituições. Como diversos autores abordam esse tema de maneira análoga, a atual pesquisa tomou como base este ponto de vista.

Os educadores podem traçar estratégias a partir da análise do perfil dos estudantes, tendo como foco deixar mais dinâmica as interações entre a relação professor-aluno, e ao reconhecer as particularidades existente pode-se personalizar o seu atendimento ao aluno (SILVEIRA *et al.*, 2015).

Ciente dessa importância, Biazus (2004) analisou várias pesquisas, na qual o mesmo citou os principais aspectos já estudados por autores que buscavam trazer à tona as causas da evasão de cursos de graduação. Entre esses aspectos ele mencionou: a identificação do sexo; estado civil; renda familiar; faixa etária e modalidade de ingresso nas IES (Intuições de Ensino Superior).

Já Stearns e Glennie (2006) apontaram que os alunos tendem a enfrentar diferentes pressões dependendo de sua etnia e gênero, e também restrições e opções

distintas de acordo com a sua da idade. Reafirmando assim, a relevância de conhecer o perfil do estudante.

A renda familiar do indivíduo é outro aspecto de grande relevância para o estudo de evasão escolar, entendendo-se que nem sempre é viável a conciliação entre o trabalho e os estudos, e pelo fato que a maioria dos cursos de línguas estrangeiras serem pagos. Em complemento, Soares *et al.* (2015) afirma que em famílias com menores condições econômicas, é mais comum que o estudante necessite trabalhar para garantir a sua subsistência ou para auxiliar nas contas da família. Ou ainda o indivíduo se sente incomodado por não poder ajudar financeiramente, ora que acredita que apenas estudar seja improdutivo.

No estudo apresentado por Neri (2010) cerca de 10,9% dos casos de evasão escolar de alunos de até 17 anos de idade é causada pela necessidade de trabalho e geração de renda. Deste modo, a colisão de horários entre curso e a atividade profissional é um fator de grande impacto no momento da decisão do aluno em continuar ou não o curso, pois geralmente os deveres profissionais se sobressaem em relação ao estudo, já que, o trabalho é responsável pela geração de renda do indivíduo.

Por sua vez, a necessidade de entrar no mercado de trabalho geralmente acarreta em uma grande carga horária. Na juventude esse fato ainda pode ser acentuado, já que os jovens querem aproveitarem a vida, e ao mesmo tempo precisam dormir por muitas horas, o que pode levar o estudante a abandonar os estudos, dado que o mesmo apresenta mais dificuldades. (ZIMMER, 2013).

Segundo Fischer *et al*. (2003) a privação de sono tende a diminuir o desempenho escolar devido aos lapsos de memória, menor concentração e menor vigilância e atenção, e devido a jornada de trabalho torna-se inviável a dedicação aos estudos fora do período escolar, aumentando o absenteísmo e atrasos nas aulas.

A pesquisa realizada pelo British Council Brasil (2014) apontou que a falta de tempo é uma notável razão para a não prática do estudo do idioma inglês. Essa pesquisa também apontou que quanto mais alta a idade da pessoa, menor é o seu tempo disponível para a realização de cursos. A falta de tempo é um aspecto de impacto negativo não somente para cursos presenciais. Corroborando com essa proposição, o estudo realizado por Araújo (2015) apontou que 80% dos alunos do

curso Licenciatura em Música EAD da Universidade de Brasília (UnB) informaram que a falta de tempo foi o principal fator causador da sua evasão.

Araújo (2015) também apontou que os alunos apresentavam dificuldades em se adaptar com as características próprias dos cursos EAD (ensino à distância), como o fato de não ter aulas presenciais, que também implicava na falta de interação com os colegas de classe e ainda ocasionava um certo desconforto de estudarem sozinhos.

Outro princípio importante é a relação entre a falta de tempo dos possíveis estudantes com a duração total dos cursos, averiguando que os cursos que aparentam ser lentos e com grande duração tendem a ser evitados, visto que cerca de 88% daqueles que planejam começar um curso de inglês esperam que o mesmo não dure mais do que dois anos. Em contrapartida, a maioria das pessoas não dão credibilidade a cursos que tenha uma duração muito breve, como aqueles que garantem ensinar conversação em apenas seis meses. (BRITISH COUNCIL BRASIL, 2014)

Outros aspectos abordados por Soares *et al.* (2015) foi que os jovens mais propícios a abandonar os estudos são aqueles que possuem baixa renda, histórico de reprovações, baixo desempenho escolar, desinteresse, falta de motivação e do sexo masculino.

Devendo-se evidenciar que a evasão escolar é uma questão complexa, que possui diversas indagações e que acontece frequentemente entre alunos. Para Freire (2014) a situação é mais agravante para aqueles indivíduos que possui idade avançada visto que a maior parte é composta por pessoas trabalhadores, casadas e com filhos.

Nesse âmbito, Borja e Martins (2014) apontaram em sua pesquisa que as questões referentes a família é um aspecto decisório em vinte e um por cento dos casos. As autoras destacaram fatores como: gravidez na adolescência; os companheiros não apoiarem suas parceiras nos estudos, condenando por estarem fora do lar nesse período; parentes que defendem a ideia de trabalhar invés de estudar, para contribuir nas despesas familiares; e ainda para se dedicar ao lar, ao esposo e filhos; e pelo fato que nem sempre os pais têm com quem deixar a suas crianças enquanto estão estudando.

Ainda na esfera familiar, Bayma-Freire *et al*. (2015) concluiu em sua pesquisa que o baixo nível de escolaridade dos pais é uma condição desmotivadora para os estudos dos filhos, já que os filhos almejam o mesmo nível de escolaridade de seus genitores, ou seja, sendo assim isto interfere diretamente na evasão escolar dos filhos. Também foi citado pelos autores a falta de incentivo dos pais, já que a escolaridade não é tão valorizada.

Além do problema da falta de incentivo, alguns sujeitos enfrentam dificuldades referentes distância da escola ou de falta de vagas próximas as suas residências. Segundo pesquisa da Galeria de Estudos e Avaliação de Iniciativas Públicas (GESTA, 2017), relata que tradicionalmente cerca de 5% dos jovens enfrentam esse tipo de problema, tornando um dos motivos relevantes para a falta de engajamento escolar. O problema de mobilidade, ainda é mais crítica para áreas rurais, devido a distância e em virtude de parte das estradas não serem pavimentadas. Já as regiões metropolitanas sofrem com as más condições do transporte, trânsito e longas distâncias, tomando parte do tempo dos estudantes, e tipicamente esse tempo gasto acaba sendo improdutivo.

Os transportes públicos para cursos de idiomas não são gratuitos, deste modo, o aluno precisa arcar com o preço do transporte. Sob esse contexto, o fator da mobilidade pode ser ainda mais relevante para alunos de cursos de idiomas.

No âmbito do aprendizado, também há diversas contrariedades que podem levar o aluno desistir. Sendo assim, a não identificação dos problemas da falta de aprendizado do sistema educacional faz que os conteúdos sejam expostos aleatoriamente â vivencia dos alunos, sendo que o ideal seria apontar as reais causas do fracasso escolar, ou seja, reconhecer o que faz o aluno a não acreditar no que faz. O fato de desconhecer os motivos existentes que geram as dificuldades de aprendizado, acarreta no aumento da repetência e da evasão escolar. (BEZERRA, 2014).

Segundo Gonçalves *et al*. (2017) o aprendizado é um ato individual que depende do seu próprio desenvolvimento intelectual, visto que de acordo com a sua vivência cada pessoa cria o seu próprio caminho de aprender, ou seja esse desenvolvimento nem sempre é simultâneo ao processo de ensino convencional.

Sousa (2008) cita o modelo de Tinto (1975) que tem como diretriz a ideia que a evasão ocorre devido à falta de integração do aluno com a classe, principalmente

nos casos de adaptação, isolamento e contradição. Enquanto Lopes (2011) cita a interação do aluno com o professor como sendo imprescindível para o êxito do processo de aprendizagem, afirmando que todo o processo de aprendizagem humana, se dá através da interação social.

O professor é o maior patrimônio de uma instituição de ensino. Sendo assim, a decepção com o professor pode ter grande influência na desmotivação do aluno. Essa decepção pode se dar por vários motivos, entre eles: falta de pontualidade, didática insuficiente, falta de planejamento ao decorrer das aulas, conhecimento insuficiente sobre a disciplina ministrada, desacordo entre a avaliação e o conteúdo estudado, falta de respeito, entre outros.

Nessa conjuntura, entende-se que a relação professor aluno é essencial para motivar bons resultados de aprendizagem, envolvendo fatores cognitivos e sócio emocionais. Segundo Fonseca (2008), isso exige competências e habilidades do professor de modo que induza os alunos ao estudo.

Em contrapartida, a educação se trata de um serviço, a aula é produzida instantaneamente enquanto o processo ocorre. Por esta razão, é inviável uma padronização por completo do ensino, pois mesmo que os materiais e a metodologia sejam homogêneos, a tendência é que cada aula tenha as suas peculiaridades, devido a maneira de cada professor ministrar o conteúdo. (SOUSA, 2008). Sendo assim, esse tipo de serviço é dependente dos alunos, sendo improvável uma aula produtiva se os alunos não possuírem conhecimentos básicos requeridos pelo curso e/ou estejam desmotivados.

Outras duas vertentes que apontam o egresso nos cursos se referem ao posicionamento do estudante a respeito do nível de domínio do conteúdo estudado, havendo dois tipos de alunos: aqueles que estavam abaixo, e aqueles que estavam muito acima do patamar compreensão do inglês. Aqueles que apresentam baixo domínio quando comparados aos demais tendem a ficarem inibidos durantes as aulas, ocasionando uma desmotivação. (MARTINS, 2011).

Como já mencionado a motivação é um aspecto preponderante para a aprendizado do aluno. Contudo, geralmente os estudantes tendem a ser mais empolgados no período inicial do curso, entretanto, ao decorrer do tempo esse entusiasmo tende a diminuir, e em alguns dos casos essa empolgação sede lugar para as decepções e frustações.

Em concordância, Robbins (2005) defende a ideia que a motivação não se refere a uma característica individual, ou seja, é errôneo dizer que alguém é desmotivado ou preguiçoso. Para ele a motivação se resulta das interações do sujeito com as situações.

A desmotivação pode ser ocasionada pela perda de interesse por parte do aluno ao perceber que criou expectativas infundadas devido à falta de informação, a não concretização e a frustação dos objetivos traçados pelo aluno ou até mesmo a falta de identificação com o curso escolhido. Esses são aspectos que contribuem para uma possível evasão, o que não se reflete necessariamente na desistência de continuar os estudos, mas uma reopção de formação. (GOMES, 1998).

Sob esse enredo, Vergara (2009) citou a Teoria da Expectativa de Victor Vroon que faz uma associação entre desempenho e recompensa, defendendo a ideia que o sujeito se sente motivado a esforçar-se em realizar algo, quando o mesmo acredita que isso lhe renderá recompensas que suprem seus objetivos individuais.

Trazendo essa teoria para o atual contexto da pesquisa, acredita-se que essa recompensa possa estar ligada ao grau de importância e relevância de estudar idioma, seja para a sua vida profissional e/ou pessoal. Contudo, no relatório de *British Council* (2014) frisou que o público geralmente opte por se comprometer com o ensino superior, especializações, e MBAs, visto que os mesmos são mais efetivos na garantia de uma futura promoção ou aumento salarial. Apenas quando há condições financeiras sobressalentes, que os indivíduos buscam realizar demais cursos, entre eles de aprendizado ou aperfeiçoamento do inglês.

Deve também salientar o nível de satisfação ou insatisfação experimentada pelo aluno influenciará seu comportamento, ou seja, um estudante satisfeito provavelmente irá realizar a rematrícula a cada ciclo e tenderá a falar bem da instituição para os demais. Por outro lado, um estudante insatisfeito, provavelmente, se irá desligará e consequentemente a falar mal . (SOUSA, 2008).

Nesse contexto, Pereira e Botelho (2009) afirmam que satisfação do aluno pela a instituição de ensino tende a minimizar as evasões, sendo um fator essencial para assegurar a estabilidade operacional e financeira, mantendo a sua viabilidade mercadológica e econômica.

Barlem *et al*. (2012) mencionou variados motivos que levam o aluno a desistir de um curso, tendo como exemplo: imaturidade, dificuldades de adaptação ao meio

acadêmico, adversidades financeiras, familiares, insuficiência ou desconhecimento de informações relacionadas ao curso, ou até mesmo a decepção com a formação escolhida.

Já para Hotza (2000) as possíveis razões para a alta taxa de evasão escolar podem estar diretamente ligadas a falhas das escolas, e a fatores psicológicos e econômicos financeiros dos estudantes. Nesse estudo, a autora apresentou exemplos como:

a) Precariedade de aparelhos e de materiais disponíveis;
b) Desleixo da universidade para com os cursos de baixo prestígio;
c) Decepção com o curso/professor;
d) Colisão de horários entre curso e a atividade profissional;
e) Modificações de interesses pessoais;
f) Longo percurso até o local das aulas;
g) Falta de aptidões para a profissão escolhida;
h) Desejo de experimentar um novo curso;
i) Por motivo de saúde.

Analisando os exemplos da autora pode-se notar a similaridade entre as possíveis causas da evasão escolar com as evasões de cursos de línguas estrangeiras.

Os exemplos das alíneas “a” e “b”, são menos recorrentes nas escolas de idioma, no entanto ainda é possível notar que instituições que oferecem mais de um idioma tendem a focar-se naquele que gera maior rentabilidade.

A fim de sintetizar os fatores apontados nessa seção como sendo possível causadores da evasão escolar, foi criada a seguinte Tabela 1:

Tabela 1 - Síntese dos aspectos que contribuem para uma possível evasão escolar e seus respectivos pesquisadores

Pesquisadores	**Aspectos Indicados**
Stearns e Glennie (2006)	Etnia e gênero.
Soares *et al. (2015)*	Baixa renda; histórico de reprovações; baixo desempenho escolar; desinteresse; falta de motivação; sexo masculino.
Neri (2010)	Necessidade de trabalho e geração de renda; colisão de horários .
Zimmer (2013)	Dificuldade de conciliar a cargas horária.
Fischer et al. (2003)	A falta de tempo tende aumentar o absenteísmo e atrasos nas aulas.

British Council pelo Instituto de Pesquisa Data Popular (2014)	Quanto mais alta a idade da pessoa, menor é o seu tempo disponível para a realização de cursos.
	Preferência por cursos universitários e MBAs, visto que são mais efetivos na garantia de uma futura promoção ou aumento salarial.
	Duração total do curso, uma vez que cursos muito longo ou lentos tendem a ser evitado, e aqueles com duração muito curta não ganham credibilidade.
Araújo (2015)	Dificuldades de adaptação com as características dos cursos EAD.
Freire (2014)	A Idade avançada, visto que a maior parte é composta por pessoas trabalhadores, casadas e com filhos.
Borja e Martins (2014)	Aspectos familiares, como gravidez na adolescência, parentes que defendem a ideia de trabalhar invés de estudar, para contribuir nas despesas familiares; e pelo fato que nem sempre os pais têm com quem deixar a suas crianças enquanto estão estudando.
Bayma-Freire et al. (2015)	Baixo nível de escolaridade dos pais é uma condição desmotivadora para os estudos dos filhos.
Gesta (2017)	Distância da escola ou de falta de vagas próximas as suas residências.
Bezerra (2014)	Falta de identificação dos problemas causadores da evasão.
Tinto (1975 *apud* Sousa, 2008)	Falta de integração do aluno com a classe.
Lopes (2011)	Importância da interação do aluno com o professor.
Martins (2011)	Nível de domínio muito abaixo ou muito acima da classe.
Gomes (1998)	Desmotivação.
Sousa (2008)	Satisfação ou insatisfação experimentada pelo aluno influenciará seu comportamento.
Barlem et al. (2012)	Imaturidade, dificuldades de adaptação ao meio acadêmico, adversidades financeiras, familiares, insuficiência ou desconhecimento de informações relacionadas ao curso, ou até mesmo a decepção com a formação escolhido.
Hotza (2000)	Falhas das escolas; fatores psicológicos e econômicos financeiros dos estudantes.

Fonte: Elaborada pela autora, 2019.

2.2. Aprendizado de máquina

O aprendizado de máquina também conhecido como *Machine Learning* (termo em inglês) vem se destacando de maneira notória, tornando se um dos pilares da tecnologia da informação. (SMOLA, 2008). Sendo considerado uma ramificação da inteligência artificial baseada no conceito de que sistemas são capazes de aprender com dados, reconhecendo padrões e adotando decisões com o mínimo de intervenção humana, ou seja, é uma metodologia de análise de dados que automatiza a criação de modelos analíticos. (CAMARGO, 2018).

Atualmente há uma série de dispositivos que fazem uso de tecnologias de Inteligência Artificial e aprendizado de máquina. Entre as aplicabilidades do

aprendizado de máquina estão: Detecção de Fraudes; Sistemas de Recomendação; Mecanismos de Busca; Sistemas de Vigilância em Vídeo; Reconhecimento de Manuscrito; Processamento de Linguagem Natural; *Bots* de Serviço ao Cliente; Segurança de Tecnologia da Informação (TI); Análise de *Streaming* de Dados; Manutenção Preditiva; Detecção de Anomalia; Previsão de Demanda; Logística; Negociação Financeira; Diagnóstico de Cuidados de Saúde; Veículos Autônomos; Robôs, entre outras. (DATA SCIENCE ACADEMY, 2018).

Os modelos de aprendizado de máquinas têm como base o conceito que o computador aprende por meio de experiências E, melhorando a sua competência em um conjunto de tarefas T, com uma medida de desempenho D. (MITCHELL, 1997).

Os autores Shalev-Shwartz e Ben-David (2014) recomendam a utilização de técnicas de aprendizado de máquinas em problemas que podem exigir além da capacidade cognitiva, devido a sua complexidade, na qual, geralmente estes dilemas possuem diversas variáveis. Outro exemplo seria atividades que podem ser realizadas por animais/ humanos, no entanto não conseguimos explicar detalhadamente de como conseguimos fazer, por exemplo, reconhecimento da fala e compreensão de imagens.

Já para Roza (2016) o aprendizado de máquina tem como objetivo resolver os problemas do mundo real que expressem importância e possuam bases de dados contendo informações que possam ser capazes de obter uma solução.

Em complemento Domingos (2015) afirma que o aprendizado não é uma mágica ou algo surpreendente, e sim o fato de obter mais do menos, como no caso da agricultura, onde a natureza é responsável pela maior parte do trabalho, enquanto os agricultores buscam combinar as sementes com os nutrientes para o plantio. Já na aprendizagem os pesquisadores combinando os dados com os conhecimentos para desenvolver seus programas.

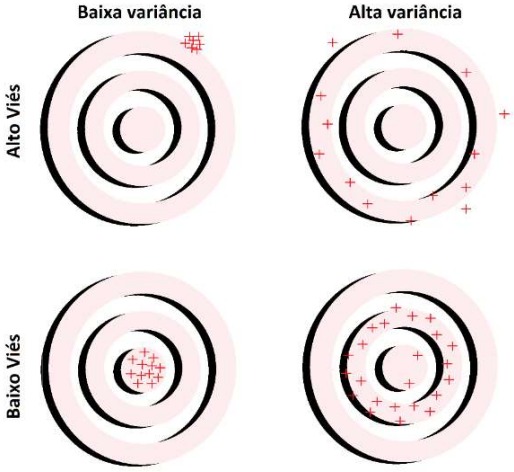
Pautados do senso comum, os humanos são capazes de reconhecer aprendizados que não lhe apresentam sentido, como se fosse uma espécie de filtro de conclusões aleatória, no entanto, quando passamos a tarefa de aprender para a máquina é necessário propiciar princípios bem definidos, que funcione como uma proteção da técnica, para que o mesmo não chegue em conclusões inúteis e/ou sem sentido. (SHALEV-SHWARTZ; BEN-DAVID, 2014).

Já os autores Shalev-Shwartz e Ben-David (2014) explanaram sobre a inevitabilidade da incorporação de conhecimentos prévios para o sucesso dos programas de aprendizado, no entanto, deve se atentar ao fato que quanto mais forte for essa incorporação de conhecimento ou de suposições prévias, mais fácil de aprender com os exemplos e menos flexível será o aprendizado.

Além do mais, é imprescindível que haja um volume de dados disponível adequado suficientemente bom para representar o conjunto como um todo. Pois caso contrário, a sua má definição poderá causar problemas na construção de modelos de aprendizado, o que pode ser agravado pelo fato que este tipo de problema só se torna visível quando inseridas novas instâncias desconhecidas no modelo. (CARVALHO, 2014).

Um dos desafios da programação de aprendizado de máquinas é encontrar uma técnica com baixa variância e baixo viés, no entanto, isto geralmente é chamado de *trade-off*, pois esses parâmetros estão em polos opostos, ou seja, quando diminui-se um deles o outro tende a aumentar, sendo fácil obter um método de variância muito baixo, mas com um alto viés ou com um viés baixo, mas com uma alta variância. (JAMES *et al*., 2015). Afim de ilustrar o caso, foi feita uma analogia com o jogo de dardos como se pode ver na Figura 1.

Figura 1 - Viés e variação no lançamento de dardos



Fonte: Adaptado de Domingos (2015, tradução do autor)

Ainda para Domingos (2015) um classificador que possui uma alta complexidade tende a ter uma variância alta, pois aumenta a possibilidade de a vir a ter conclusões irreais para que o mesmo seja capaz de apreender com diversos

padrões. Isso se dá pelo fato de que a variância é uma a tendência de se aprender eventos aleatórios independentemente do real. Já o viés pode ser definido como uma tendência de um classificador aprender coerentemente uma generalização errônea.

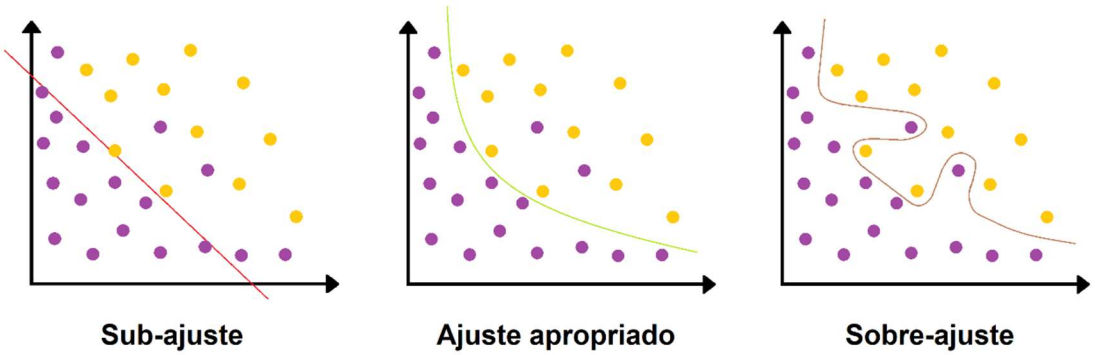
Jabbar e Khan (2015) em sua pesquisa também destacou que o sobre-ajuste (do inglês: *overfitting*) é um problema comum em tarefas de aprendizado de máquina supervisionado. Os autores explicaram que o sobre-ajuste é um fenômeno que ocorre quando o modelo matemático se adapta tão bem aos dados de treino que quando é testado em um conjunto de dados desconhecido, o seu desempenho cai, resultando em um desempenho não confiável quando praticado em novas medições, de modo que esse fenômeno é mais frequente em bancos de dados pequenos.

Carvalho (2014) deu como exemplo de sobre-ajuste de dados de treino, um modelo que possui uma performance de 100% de precisão em um conjunto de dados de treinamento, porém apresenta uma precisão de apenas 50% nas instâncias de teste

Evitar o sobre-ajuste pode parecer contraditório, posto que o proposto das técnicas de otimização é alcançar a melhor solução possível no espaço de parâmetros conforme pré-definida a função objetiva e dados disponíveis. (JABBAR; KHAN, 2015).

Já o sub-ajuste (do inglês: *underfitting*) é o oposto do sobre-ajuste, ou seja, o modelo é incapaz de captar a variabilidade dos dados, não sendo capaz de se ajustar bem aos dados de treinamento. Isso geralmente é atribuído ao fato de usar um modelo muito simples para a previsão. (DIAS *et al*., 2016). Para ilustrar o sobre-ajuste e o sub-ajuste foi criada a Figura 2.

Figura 2 - Ilustração de Sobre-ajuste, Sub-ajuste



Fonte: Adaptado de Radiya-Dixit (2017, p. 2, tradução do autor)

O aprendizado de máquina é subdividido em quatro grandes grupos de acordo com as suas características e técnicas particulares: supervisionado, semi-supervisionado, não supervisionado e por reforço.

O objetivo do aprendizado supervisionado é induzir conceitos a partir de exemplos que já estão pré-classificados. Pode ser categorizado como classificação se as classes possuírem saídas categóricas (rótulos). Entretanto se as classes possuem saídas numéricas, define-se como um problema de regressão. (DIAS *et al*., 2016).

Na aprendizagem não supervisionada a técnica funciona sem um supervisor, não existindo uma influência direta humana e o aprendizado se baseia em um agrupamento natural a partir da similaridade dos seus atributos (NEVES, 2018). Enquanto o semi-supervisionado é uma combinação do aprendizado supervisionado e não supervisionado, é utilizado esse método quando a quantidade de dados rotulados é baixa, usando assim dados não rotulados para compor o conjunto de treinamento. (DIAS *et al*., 2016).

O aprendizado por reforço é um sistema que recebe *feedback* análogo a punições e recompensas, conseguindo aprender a partir de experiências do mesmo com algum meio. (DATA SCIENCE ACADEMY, 2018). Geralmente esse tipo de técnica é usada em jogos e na robótica. Na Figura 3 é mostra a hierarquia do aprendizado indutivo.

Figura 3 - Hierarquia do aprendizado de máquina supervisionado e não supervisionado

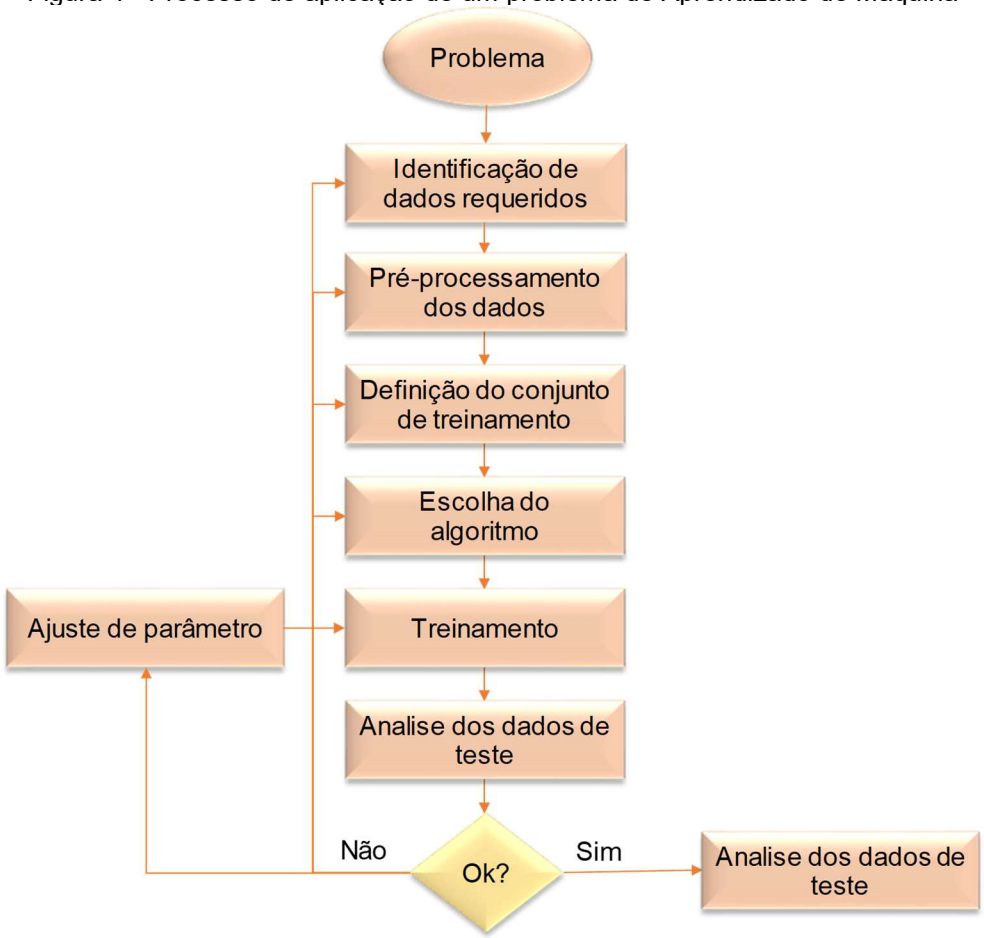
Fonte: Adaptado de Monteiro (2018, p. 16).

Nos próximos subtópicos será discutido com maior nível de detalhamento o aprendizado supervisionado e os não supervisionado pois estes representam a grande maioria das técnicas de aprendizado de máquina e pelo fato que foram usadas apenas técnicas referentes a estes grupos.

2.2.1. Aprendizado de máquina supervisionado

A Aprendizagem Supervisionado, também comumente chamada de classificador, consiste em usar uma série de instâncias previamente classificados, afim de que a técnica seja capaz de aprender com esse rótulo, e posteriormente consigo classificar de maneira precisa um novo conjunto de dados. (CARVALHO, 2014). Tem como objetivo construir um modelo conciso para fazer previsões sobre casos futuros, em que é necessário realizar uma série de processos para obter tais resultados, conforme esquematizado na Figura 4. (KOTSIANTIS, 2007).

Figura 4 - Processo de aplicação de um problema de Aprendizado de Máquina



Fonte: Adaptado de Kotsiantis (2007, p.250, tradução do autor)

Segundo os autores Mohammed e Pathan (2013) a máquina é capaz de aprender e de perceber os *feedbacks* de seu ambiente.

2.2.2. Aprendizado de máquina não supervisionado

A aprendizagem não supervisionada geralmente busca descobrir padrões ocultos ou detectar anomalias nos dados, ou seja, encontrar categorias que seriam considerados apenas ruídos não estruturados. Isso é possível através do desenvolvimento de uma estrutura formal, tendo como princípio a meta de que a máquina precisa criar representações de entrada que podem ser usadas na tomada de decisões, tendo como propósito a previsão de entradas futuras. (MOHAMMED; PATHAN, 2013).

O aprendizado não supervisionado em geral é mais desafiador do que o aprendizado supervisionado, dado que o mesmo tende a ser mais subjetivo, deste modo, é usualmente mais utilizado como parte de uma análise de dados exploratória. Além do que, pode ser mais difícil a avaliação dos resultados alcançados, uma vez que não há mecanismo universalmente aceito para a realização da validação cruzada ou a dos resultados obtidos em um conjunto de dados independente, já que não se sabe as respostas verdadeiras dos dados. (JAMES *et al*., 2015).

2.3. Aplicação do aprendizado de máquina na evasão escolar

O trabalho de Hotza (2000) ressaltou que evasão não pode ser explicada por umas simples respostas, e sim por um conjunto de fatores inter-relacionados, justificando-se a utilização de técnicas de aprendizagem de máquina para este tema.

Enquanto Aguiar (2018) realizou um estudo sobre a previsão da evasão de aluno em turmas de Introdução à Programação de Computadores (IPC) que utilizam juízes online por meio da utilização da técnica árvore de decisão que possibilitou uma acuracidade de 80% na predição de evasão de alunos.

Sob esse contexto, Kantorski *et al.* (2015) propôs realizar uma análise das informações pessoais, acadêmicas, de caráter sócio econômico dos estudantes do curso de Zootecnia, combinando essas informações para criar um modelo preditivo de evasão com o auxílio da mineração de dados e de técnicas de aprendizado de

máquina, alcançando 98% de acuracidade na previsão das classes envolvidas (aluno regular e aluno evadido).

Já Pinheiro *et al.* (2018) sugeriu em sua pesquisa utilizar as técnicas *Naive Bayes* (NB), Máquina de Vetores de Suporte e Árvore de Decisão para a previsão de evasão no Ensino Superior concluindo que os mesmos conseguem prever adequadamente os casos.

Ainda, no trabalho de Brito *et al.* (2014) apresentou bons resultados ao estudar as técnicas Métodos Bayesianos, Métodos de Vizinho mais Próximo, Máquina de Vetor de Suporte, Árvore de decisão e Redes Neurais, como meio de compreender a relação entre o desempenho do estudante do primeiro período com a nota de ingresso na universidade.

2.4. Técnicas de aprendizado de máquina

2.4.1. Análise de componentes principais

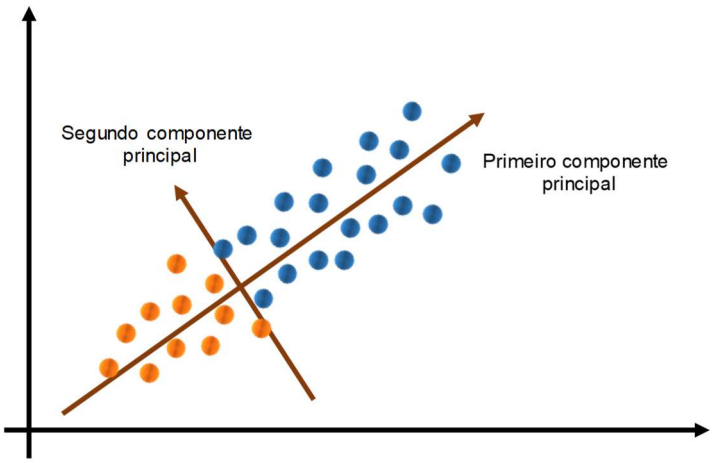
A Análise de Componentes Principais (PCA, do terno inglês: *principal component analysis*) é um método multivariado de modelagem, que possui como proposito transformar linearmente um conjunto de elementos, inicialmente correlacionadas entre si, em um conjunto significativamente menor de variáveis não correlacionadas, no entanto que possuem a maior parcela de informações do conjunto original. (HONGYU *et al*., 2015).

Segundo James *et al*. (2015) o PCA é um método de aprendizado não supervisionado usado para a visualização ou pré-processamento dos dados antes da aplicação de técnicas de aprendizado supervisionado.

A ideia principal do PCA é utilizar um sistema de coordenadas especial dependente da distribuição dos pontos no espaço da seguinte forma: posicione o primeiro eixo na direção que possui uma maior variação entre pontos afim de maximizar a variação ao longo desse eixo, em seguida posicione o segundo eixo perpendicular ao primeiro. Deve-se frisar que em duas dimensões não há como ter escolha da direção do segundo eixo, uma vez que a mesma é definida de acordo com o primeiro, no entanto em mais dimensões, os demais eixos podem apontar diversas direções. É aconselhável escolher a posição do segundo eixo de maneira que

maximize a variação ao longo dele, e assim como para a definição dos demais. (WITTEN; FRANK, 2005). Para exemplificar, foi plotada a figura 5.

Figura 5 - Exemplo de um problema de projetar pontos 2D em uma dimensão



Fonte: Adaptado de Chan *et al*. (2013)

Geralmente é analisado o gráfico conhecido por "*scree plot*" (gráfico de segmento de linhas simples que exibe o percentual de variação total nos dados) como sendo um meio de decidir o número de componentes principais que serão necessários para visualizar os dados. Isso é feito através da observação do *scree plot*, buscando o ponto na qual a proporção de variância explicada por cada componente principal subsequente diminua, formando o chamado "cotovelo do *scree plot"*. Contudo, infelizmente, não há maneira objetiva bem-conceituada para decidir quantos componentes principais são suficientes. (JAMES *et al*. 2015).

Ainda segundo James *et al*. (2015), na pratica, geralmente examina-se os primeiros componentes principais na busca de padrões que sejam interessantes para o estudo. E caso este seja encontrado, é estendida a análise para os componentes subsequente até o instante que não seja achado nenhum outro padrão interessante. Entretanto, caso não seja encontrado nenhum padrão vantajoso desde dos primeiros componentes principais é improvável que os demais possuam algum padrão relevante. Sendo esta, uma abordagem subjetiva que reflete o fato de que a técnica PCA é comumente utilizada como uma ferramenta para análise exploratória de dados.

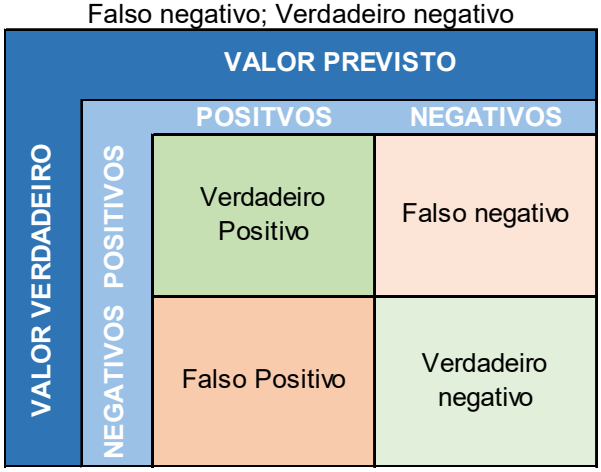
A definição de qual técnica de aprendizagem que deverá ser utilizado é um dos fatores críticos da problemática. Geralmente a avaliação do mesmo é feita com

base na acuracidade da previsão, sendo o percentual de previsões corretas, dividido pelo número total de previsões. (KOTSIANTIS, 2007).

Os resultados das técnicas podem ser plotados em uma matriz de confusão, na qual a diagonal principal da matriz apresenta a quantidade de acerto, e a soma dos demais algarismo significa a quantidade de erros. (TARCA *et al*., 2007).

A taxa de acerto é decomposta em duas vertentes: verdadeiro positivo que ocorre quando o previsto é uma variável positiva assim como os dados observados; verdadeiro negativo se dá quando ambas os valores são negativos. Enquanto a taxa de erro é composta por: falso positivo quando se prevê uma variável como sendo positiva, mas na verdade a mesma é negativa; e falso negativo, quando se classifica o elemento como negativo, mas os dados observados apontam que a mesma é positiva. (DIAS *et al*., 2016). A Figura 6 mostra a situação relatada.

Figura 6 - Matriz de confusão separada em quatro quadrantes: Verdadeiro positivo; Falso positivo; Falso negativo; Verdadeiro negativo



Fonte: Adaptado de Faceli (2011, p. 164, tradução do autor)

Também encontramos na literatura o Falso Positivo ser denominado como Erro do Tipo 1, entanto o Falso Negativo de Erro do Tipo 2.

2.4.2. K-vizinhos mais próximos

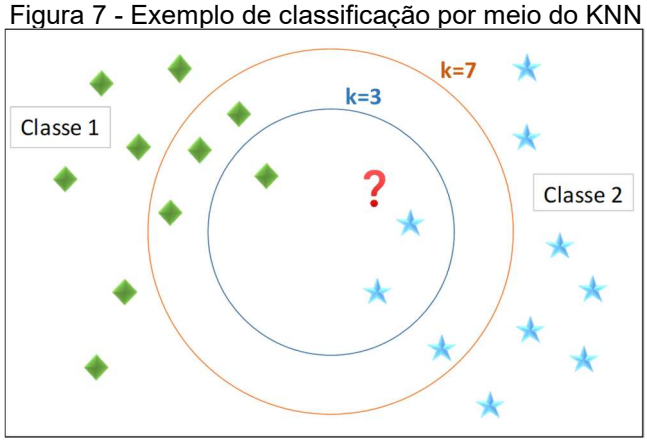
A técnica k-vizinhos mais próximos (KNN, tradução livre do termo inglês: *K-nearest neighbors*) trata-se de um método usado como regressor ou classificador em reconhecimentos de padrões. (MONTEIRO, 2018). Esse modelo usa como métrica a distância entre os pontos, deste modo, a técnica atua apenas com valores numéricos, fazendo-se necessário a transformação de características categóricas em

quantitativas. Essas características ainda precisam passar por um processo de normalização, que geralmente consiste em modificar a variação de todos os elementos para uma variação única de 0 a 1. (CARVALHO, 2014).

A técnica KNN é um procedimento de decisão não paramétrico mais simples, este consiste em atribuir a observação não classificada a classe da amostra mais próxima (usando métrica) no conjunto de treinamento. (COVER; HART, 2018).

De acordo com a regra dos vizinhos mais próximos, quando definimos K números vizinhos igual a 1, o mesmo classifica a variável como pertencente à categoria do seu vizinho mais próximo e ignora todos os outros. Já quando definimos K números vizinhos como sendo maior que 1, ele classifica a variável de acordo com a categoria referente a maioria dos elementos mais próximos. (COVER; HART, 2018). Para ilustrar, foi criada a Figura 7, tendo duas classes, uma de estrelas azuis e a outra de losango verde, na qual a chegada de uma nova variável é simbolizada pela interrogação, e a mesma precisa ser classificada.

Figura 7 - Exemplo de classificação por meio do KNN



Fonte: Adaptado de Cover e Hart (2018)

No exemplo ilustrado na Figura 7, o ponto de interrogação seria classificado em:

- K números vizinhos igual a 1: Estrela Azul (classe 2)
- K números vizinhos igual a 3: Estrela Azul (classe 2)
- K números vizinhos igual a 7: Losango Verde (classe 1)

Embora o aprendizado seja eficaz e simples, é aconselhado a sua utilização em bases de dados com muitas instâncias, pois dependendo da escolha de K números

vizinhos, o mesmo fica suscetível a ruídos quando aplicado em pequenas bases de dados. (CARVALHO, 2014).

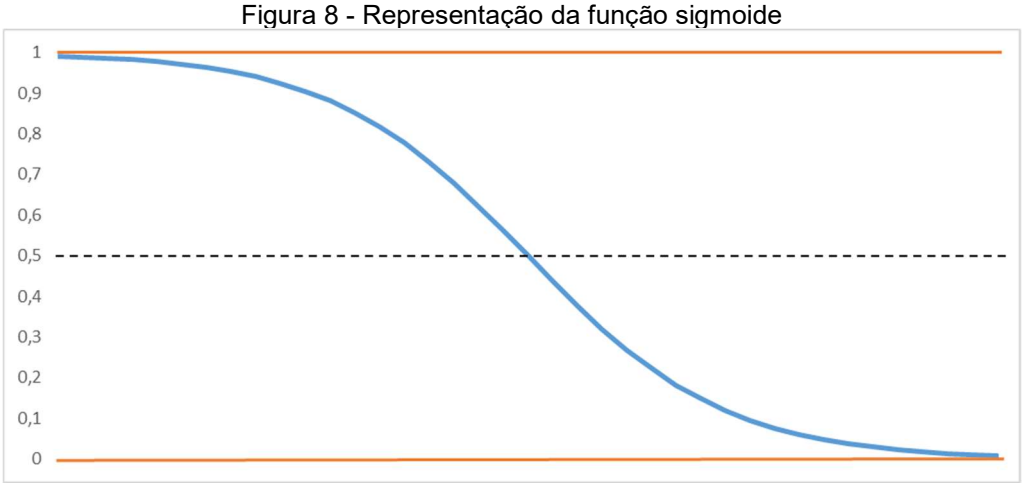
A escolha do número k de vizinhos pode afetar de maneira significativa a precisão de um modelo, sendo assim o mesmo deve ser escolhido com uma devida cautela. Se escolhermos um k muito grande, o efeito do fator mais próximo diminuíra, já que a nova instância será classificada dependente de um sistema como um todo. Um exemplo seria se a escolha de k fosse igual o número de amostra, sempre que houvesse uma nova variável, a mesma seria rotulada como pertencente a maior classe. (COVER; HART, 2018).

2.4.3. Regressão logística

A Regressão Logística (do termo inglês: *Logistic Regression*) é uma técnica estatística que foi agregada nos modelos de aprendizado de máquina afim de realizar classificações. Por sua vez a regressão logística muitas das vezes é comparada com a regressão linear, no entanto em vez de arriscar valores de probabilidade ilegítimos por meio de aproximação dos valores diretamente para 0 ou 1, a regressão logística constrói um modelo linear quando esse limite é ultrapassado, tendo como base uma variável de destino transformada. (WITTEN; FRANK, 2005).

A regressão Logística trabalha com dados categóricos e comumente com dados binários, diferindo-se da regressão linear que usa dados contínuos. Para isso a mesma é associada uma função sigmoide φsig: R → [0, 1] sobre a classe de funções lineares, como representado abaixo: (SHALEV-SHWARTZ; BEN-DAVID, 2014).

Figura 8 - Representação da função sigmoide



Fonte: Elaborada pela autora, 2019

Essa curva recebe o nome de "sigmóide" devido ao seu formato em “S", correspondendo ao enredo desta função, como mostrado na Figura 8.

Para utilizar a regressão logística como classificador de várias classes deve-se fazer uma regressão para cada uma das classes de maneira independentemente. Ou ainda pode-se abordar problemas de multiclasse que é também conhecida como classificação em pares. Para tal, é construído um classificador para cada par de classes, utilizando apenas as instâncias dessas duas classes. A saída de novo elemento é baseada de acordo com a classe mais votada. Este método habitualmente gera resultados precisos em termos de erro de classificação. (WITTEN; FRANK, 2005).

Os autores Hosmer e Lemeshow (2000) supuseram uma situação onde há 3 variáveis independente referente a raça: “Branco”, “Negro” e “Outros”, a fim de exemplificar uma classificação com mais de duas classes usando a Regressão Logística. Deste modo serão necessárias dois modelos de variáveis, sendo representado por X_1 e X_2 tendo valores binários, como mostrado na Tabela 2.

Tabela 2 - Um exemplo da codificação das variáveis de raça, codificadas em três níveis

RACE	Variáveis	
	X_1	X_2
Branco	0	0
Negro	1	0
Outros	0	1

Fonte: Adaptado de Hosmer e Lemeshow (2000)

Ainda para Witten e Frank (2005) a regressão logística tente a gerar estimativas de probabilidade mais precisas que acarretam em classificações mais assertivas que a da Regressão Linear.

2.4.4. Máquina de vetores de suporte

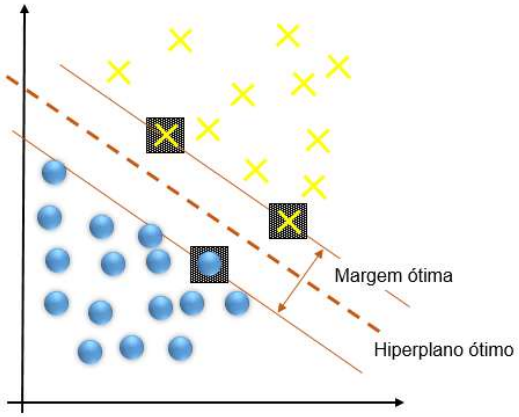
Uma máquina de vetores de suporte (SVM, do termo inglês: *support vector machine*) é um método de aprendizado supervisionado, que tem como preceito o reconhecimento de padrões através de um classificador linear binário não probabilístico. (BRITTO NETO, 2016).

Segundo Cortes e Vapnik (1995) uma máquina de vetores de suporte tem como conceito de que os vetores de entrada são mapeados de maneira não linear em um espaço característico, possuindo propriedades de decisão que garantem a sua alta capacidade de generalização de aprendizado de máquina.

O modelo SVM seleciona um pequeno número de instâncias situadas no limite crítico de cada classe, sendo denominados de vetores de suporte. Através dessa seleção é criado uma função discriminante linear, também chamada de hiperplano, que tem como objetivo separar de maneira mais ampla possível. Esse tipo de sistema possibilita a formação de limites de decisão quadrática, cúbica e até mesmo de ordens superiores, tornando-o um modelo prático para a inclusão de funções não lineares. (WITTEN; FRANK, 2005).

Para Witten e Frank (2005) o hiperplano de margem máxima é o qual proporciona uma maior partição entre as categorias. Ou seja, os elementos são mapeados de maneira que as classes são separadas com a maior distância possível entre elas. Um exemplo é mostrado na Figura 9, em que as classes são representadas por círculos azuis e “x” amarelos.

Figura 9 - Exemplo de um problema bidimensional separável, na qual as instâncias mais próximas do hiperplano de margem máxima, denominadas de vetores de suporte foram marcados com quadrados quadriculados



Fonte: Adaptado de Cortes e Vapnik (1995, p. 275, tradução do autor)

De maneira geral, o modelo SVM é caracterizado por apresentar uma boa capacidade de generalização, trabalhando de maneira robusta diante de bancos de dados de grande dimensão e pelo fato de ter apenas um único mínimo global. No entanto, as suas principais desvantagens são em relação a sua sensibilidade quanto

a definição dos valores dos parâmetros e a complexidade de interpretação do modelo obtido por essa técnica. (FACELI *et al*., 2011).

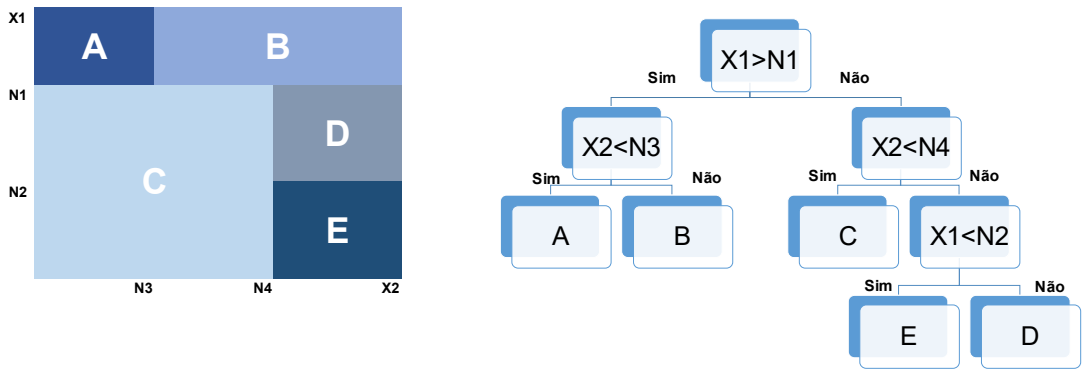
2.4.5. Árvore de decisão

A Árvore de Decisão (do termo: *decision trees*) pertence à classe de técnicas de aprendizado supervisionado, na qual pode ser aplicada em problemas de regressão e classificação (depende de qual algoritmo adotado). No entanto em nosso estudo será dado ênfase apenas na sua aplicabilidade em problemas de classificação, uma vez que esta foi usada na pesquisa.

O modelo de árvore de decisão possui vantagens como: possuir suporte para características categóricas e numéricas, não sendo necessário nenhum tipo de processamento; tem uma representação de fácil compreensão dos conhecimentos aprendidos; o treinamento dos dados é rápido quando comparado com demais modelos como o de Redes Neurais Artificiais; entre outras. (CARVALHO, 2014).

Este modelo possui uma estrutura no formato de uma árvore de decisão, onde cada nó envolve o teste de um atributo específico. Geralmente, este teste consiste em comparar um atributo com o valor de uma constante, no entanto algumas árvores comparam duas caracterizas entre si, ou utilizam alguma função de um ou mais atributos. Em contrapartida os nós de folhas fornecem a classificação obtidas a partir da aplicação de todas as instâncias que atingiam a folha, ou um grupo de classificações, ou uma distribuição de probabilidade sobre todas as classificações possíveis. (WITTEN; FRANK, 2005). Afim de exemplificar, foi criada a Figura 10.

Figura 10 - Exemplo de aplicação do método árvore de decisão para a classificação de regiões



Fonte: Adaptado de Faceli *et al*. (2011, p. 84).

Para classificação de uma nova instância, a mesma é roteada pela árvore de acordo com os valores dos atributos testados nos nós de maneira sucessiva e, quando uma folha é atingida, a instância é classificada em concordância com a classe atribuída à folha.

Segundo Faceli *et al*. (2011) o processo de poda é uma parte importante para a construção de um modelo de árvore de decisão em domínios com ruídos, uma vez que esses ruídos levantam para problemas como: a indução de classificação de novos objetos em um modo não confiável, já que nós mais profundos possuem níveis de menor importância, levando ao sobre-ajuste dos dados de treinamento; a árvore de decisão tende a ficar grande, logo, é mais difícil para compreender. No entanto poda certamente irá levar a classificação incorreta de alguns exemplos do conjunto de treinamento.

Ainda para Faceli *et al.* (2011) a técnica de árvore de decisão é o mais usado para aplicações acadêmicas, destacando a flexibilidade, robustez, seleção de atributos, interpretabilidade e a eficiência, como as suas principais vantagens.

Já para James *et al*. (2015) citou como vantagem o fato que algumas pessoas acreditam que as árvores de decisão é um das técnicas que mais se aproxima da metodologia usada para a tomada de decisão humana. Mas normalmente, as árvores não possuem o mesmo grau de acuracidade preditiva que outras abordagens de regressão e classificação, e que uma pequena variação nos dados pode gerar uma grande mudança na árvore e consequentemente na estimativa final.

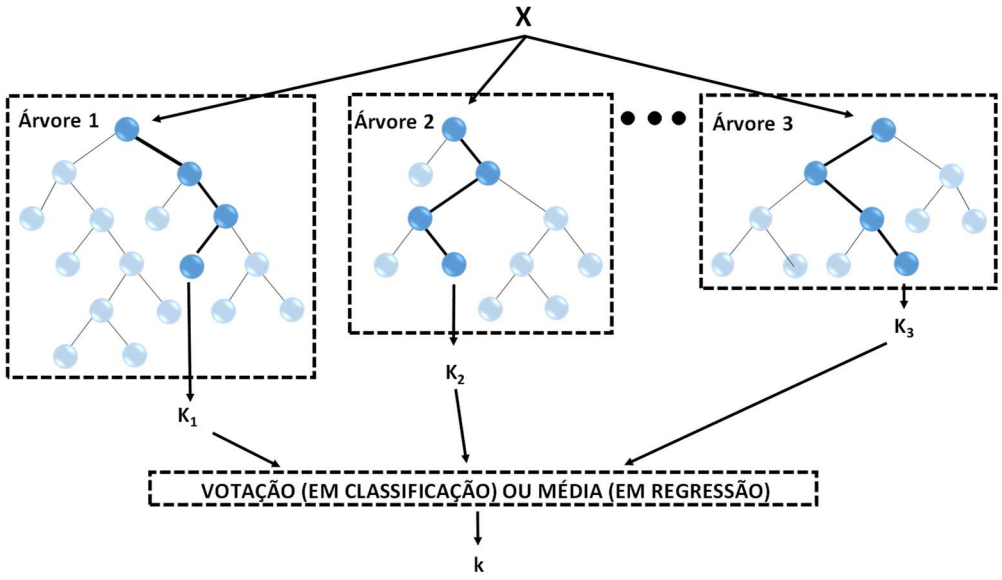
No entanto, pode-se agregar muitas árvores de decisão, usando métodos como *Bagging*, florestas aleatórias, podendo aumentar o desempenho preditivo de maneira substancialmente melhor.

2.4.6. Florestas aleatórias

Segundo Breiman (2001) as Florestas aleatórias (tradução livre do terno inglês: *random forests*) são uma junção de preditores de árvores, de modo que cada árvore é dependente dos valores de um vetor aleatório exibido de forma autônoma e com uma distribuição igualitária para todas as árvores na floresta.

Uma floresta aleatória é uma técnica de aprendizado supervisionado que pode ser usado como classificador, o mesmo se trata de um conjunto de árvores de decisão, em que cada árvore é gerada por meio da aplicação de uma técnica em um conjunto de dados de treinamento e com um vetor aleatório adicional. A classificação aleatória é tida por meio da votação majoritária sobre as previsões individuais de cada árvore, conforme ilustrado na Figura 11.

Figura 11 - Exemplo de classificação utilizando o modelo de Florestas Aleatórias



Fonte: Adaptado de Verikas *et al*. (2016, p.10, tradução do autor).

Segundo Donges (2018) em publicação no blog *Machine Learning Blog,* a floresta aleatória busca a melhor característica para fazer a subdivisão dos nós, ele procura o melhor atributo em um subconjunto randômico de características. Deste modo, esse processo é responsável pela criação de uma grande parte da diversidade do problema, o que comumente leva a geração de modelos mais eficientes.

No entanto, ainda existe um erro de generalização de classificação do modelo, que pode ser atribuído à força das árvores individuais geradas na floresta e da correlação entre elas. (BREIMAN, 2001).

Ainda Breiman (2001) afirma que as Florestas aleatórias são uma ferramenta eficaz na previsão, uma vez a incrementarão de um certo nível de aleatoriedade no modelo faz dele um classificador e regressor mais preciso.

2.4.7. Redes neurais

As Redes Neurais Artificiais (RNAs, tradução livre do termo: *Artificial Neural Networks*) é um modelo computacional inspirado na estrutura das redes biológica do cérebro, porém de um modo mais simplista. São baseadas na criação de numerosos dispositivos computacionais básicos, chamados de neurônios, estando conectados uns aos outros em uma rede de comunicação complexa, o que permite que o cérebro consiga computar dados de alta complexidade. (SHALEV-SHWARTZ; BEN-DAVID, 2014).

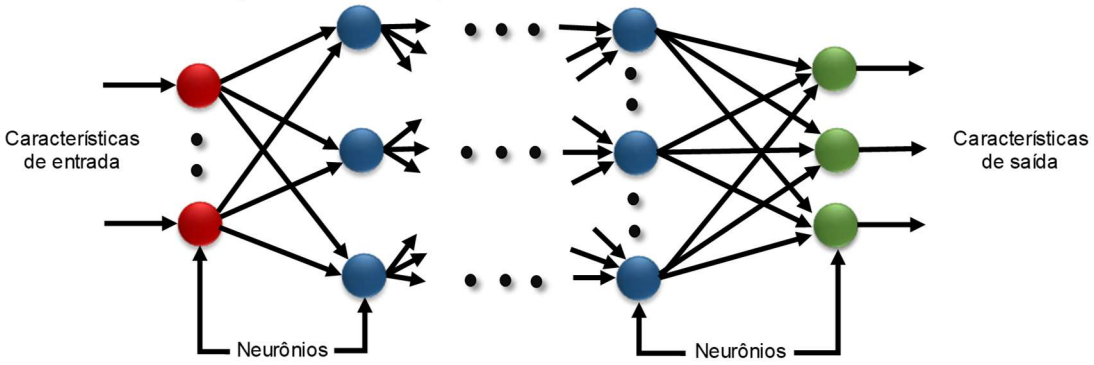
Quando analisamos tarefas como andar, pegar um objeto, identificar imagens, comumente as classificamos como simples. No entanto, quando tentamos ensina-las para um robô fica evidenciada a sua complexidade, de modo que só somos capazes de realiza-las graças a nossa estrutura biológica, onde o cérebro humano é visto como um dos principais agentes. (FACELI *et al*., 2011).

Visto a capacidade de processamento do nosso sistema nervoso, as Redes Neurais Artificiais tomaram tal como inspiração, tendo como objetivo simular essa capacidade de aprendizado e aquisição de conhecimento. No entanto o cérebro de um humano possui em torno de 10 a 500 bilhões de neurônios e um tempo de execução na ordem de 10^{-3} segundos , ou seja é diversas vezes mais rápido que um computador digital, evidenciando assim a sua superioridade em relação as redes neurais artificiais. (FACELI *et al*., 2011).

As RNAs possuem duas vertentes básicas: a arquitetura diz a respeito quanto ao tipo e número de unidades de processamento; e o aprendizado está relacionamento as regras de ajustes dos pesos de rede e que informação será usada por tais regras. (BRAGA *et al*., 2007).

Neste modelo tem-se que os dados são incorporados na camada de entrada e posteriormente esses dados são multiplicados pelos pesos aleatoriamente gerados. Enquanto as saídas são definidas através de uma função de ativação e entrada total, podendo assumir formato de função linear, limiar e sigmoidal, onde os neurônios podem estar arranjados em uma ou mais camadas. Quando se tem um sistema com mais de uma camada, o denominasse de rede multicamadas, nessa conjuntura um neurônio recebe como entrada os valores de saída da camada que a antecede, como ilustrado na Figura 12. (FACELI *et al*., 2011).

Figura 12 - Exemplo de uma Rede de Multicamadas



Fonte: Adaptado de Ebrahimi *et al*. (2012, p. 13, tradução do autor)

As entradas da rede são apresentadas na primeira camada, que é denominada camada de entrada, representada pela cor vermelha. A camada de entrada é responsável por distribui as informações para a(s) camada(s) escondida(s) da rede, por sua vez está indicada na cor azul. A última camada é a camada de saída (cor verde), onde é obtida a solução do problema. (SANTOS *et al.*, 2005).

Para Santos *et al*. (2005) é necessário escolher com cautela os parâmetros de implementação que interferem no desempenho das RNAs, citando os seguintes paramentos:

- Número de nós na camada de entrada, ou seja, a quantidade de atributos que serão utilizadas como fonte de alimentação da RNA (geralmente se determina de acordo com as variáveis mais relevantes);
- Número de neurônios e camadas escondidas que serão empregues, teoricamente uma camada já seria o bastante, enquanto o número de neurônios ocultos é determinado por meio de critérios de ajustamento-penalidade, devendo ressaltar que poucos neurônios escondidos levam o modelo ter uma generalização, mas podem gerar um sub-ajuste, ou seja o modelo não converge;
- Número de neurônios na camada de saída.

As redes que possuem apenas uma camada são capazes apenas de classificar variáveis que sejam linearmente separáveis, enquanto a rede multicamadas conseguem trabalhar com qualquer função contínua. (BRAGA *et al*., 2007).

As RNAs como citado anteriormente se inspiram nas redes biológicas, porem do ponto de vista físicas, ambas se diferenciam de maneira significativa uma da outra,

apesar que possuírem similaridades em comum, como por exemplo o fato que os dois sistemas são baseados em unidades de computação paralela e distribuídas de modo que se comunicam através de conexões sinápticas, possuindo detectores de atributos, redundância e modularização das conexões. Mesmo com a pouca semelhança, esses traços em comum entre os sistemas permitem que as RNAs sejam capazes de produzir algumas funções que antes apenas os humanos conseguiam desempenhar. (BRAGA el al., 2007).

Ainda segundo Faceli *et al*. (2011) o entendimento de como é feita a tomada de decisões das RNAs são complexos, uma vez que que os seus parâmetros estão ligados por meio de complicadas fórmulas matemáticas.

3. MÉTODO DE PESQUISA

Nesse capítulo serão abordados de maneira sintetizada os passos adotados para o desenvolvimento desta pesquisa, como a estrutura do instrumento de coleta de dados, o período e a metodologia para a aplicação dos questionários, assim como descrição dos meios utilizados para implementação das técnicas de aprendizado de máquina.

3.1. Etapas da Pesquisa

Primeira foi identificado que a evasão é uma problemática relevante para os cursos de línguas estrangeira, e então verificou-se que o aprendizado de máquina poderia ser usado como uma ferramenta no combate da evasão por meio de técnicas de predição. Sendo assim foi levantado o referencial teórico das possíveis causas da evasão escolar, assim como trabalhos correlatos do tema, e posteriormente realizou um estudo sobre o aprendizado de máquina e suas metodologias, em que foram selecionadas as técnicas K-vizinhos mais próximos, Regressão Logística, Maquina de Vetores de Suporte, Árvore de Decisão, Florestas Aleatórias, Análise de Componentes Principais e Redes Neurais para serem implementadas.

Após as pesquisas foi elaborado um questionário, partindo do princípio que esse fenômeno pode se dar tanto por motivos particulares como por fatores externos. Em seguida os questionários foram aplicados em uma pequena amostra escolhida por conveniência, afim de identificar possíveis falhas. Essa amostra era composta por professores, alunos de graduação, e indivíduos com o ensino médio completo

A partir dessas amostras foi realizado alguns ajustes segundo a percepção da autora levando em consideração apontamentos apresentados pelos respondentes. Após realizar todas retificações, os dados amostrais foram deletados e os formulários foram efetivamente aplicados.

Para iniciar a pesquisa o participante precisava aceitar o Termo de Consentimento Livre e Esclarecido, caso contrário a pesquisa se dava por encerrada. Esse termo tinha como propósito informar o participante de modo que este pudesse manifestar de maneira consciente e autônoma.

Na fase de pré-processamento os dados categóricos foram transformados em quantitativos, como por exemplo, quando a pessoa declarava que era do sexo masculino essa resposta era alterada para o número “1” e “2” para os indivíduos do sexo feminino e o número “3” para aqueles que prefiram não se identificar. A ordem da numeração dada era definida de acordo com a ordem cronológica das respostas, ou seja, como a primeira pessoa a responder foi um homem, logo o sexo masculino era intitulado como sendo o número “1”, e assim sucessivamente. Além disso, foram excluídos dados considerados irrelevantes para essa parte do estudo, como horário e dia das respostas, termos de aceite, e-mails, sugestões e *feedbacks.*

Ressalta-se que nessa fase de pré-processamento não foi necessário nenhum tipo de tratamento para dados ausentes, uma vez que as perguntas consideradas relevantes para pesquisa requisitavam respostas, ou seja, o respondente não conseguia avançar no questionário enquanto estas não fossem respondidas.

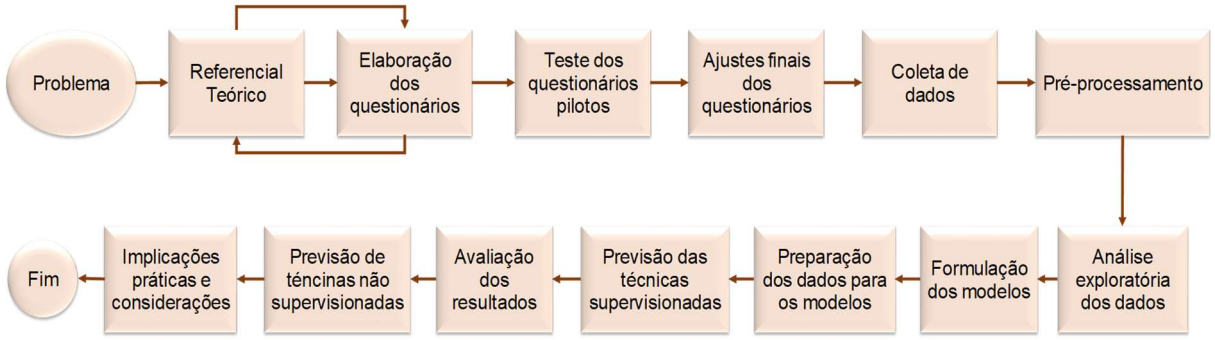
Então, realizou-se uma análise exploratória dos dados, buscando caracterizar os participantes da pesquisa. Sucessivamente para a implementação das técnicas foram traçadas dois modelos e duas configurações distintas, dado que o primeiro modelo buscava prever se o aluno estava cursando, se havia evadido ou se havia concluído o curso de idioma, enquanto o segundo modelo era binário, buscando prever se o aluno vai desistir ou não do curso.

Na preparação dos dados para o modelo, foi realizada a divisão aleatória em dois conjuntos, sendo que 70% dos dados foram destinados para treinamento, tendo como objetivo proporcionar o aprendizado das técnicas. E os outros 30% dos dados foram usados após a implementação da técnica de aprendizado de máquina para testar a acuracidade do modelo, na qual era solicitado que a técnica classificasse estes dados de acordo com a sua metodologia, e em seguida era comparado as respostas preditas pela técnica versus as respostas verdadeiras.

Foi avaliado e comparado os resultados dos modelos assim como os desempenhos individuais das técnicas, sendo selecionadas duas técnicas do Modelo 2 para a aplicação do aprendizado não supervisionado, tendo como propósito prever se os alunos que estão atualmente cursando iram concluir ou não o curso. E por fim, discutiu-se as conclusões obtidas sobre as implicações práticas da atual pesquisa e

suas contribuições acadêmica. A fim de representar as etapas dos estudos, foi criada a Figura 13:

Figura 13 - Fluxograma das etapas da pesquisa



Fonte: Elaborada pela autora, 2019

3.2. Coleta de dados e estrutura do questionário

O levantamento de dados foi feito por meio da aplicação de questionários *online* usando a ferramenta *Google Forms*, que ocorreu entre o dia 19 (dezenove) de fevereiro à 26 (vinte e seis) de abril de 2019 (dois mil e dezenove). A divulgação ocorreu por meio de redes sociais, e-mail institucional para os alunos de Engenharia de Produção da UTFPR – câmpus Londrina, comunicados impresso com o QR-Code do formulário e por e-mail para rede de contato da pesquisadora assim como para potenciais respondentes.

O questionário obteve uma taxa de 341 respostas, na qual 32 pessoas informaram que nunca fizeram curso de idioma, visto que essa classe não faz parte da área interesse deste trabalho, as mesmas foram desconsideradas, sendo assim, foram analisadas 309 respostas.

A pesquisa tinha como fator chave a divisão de 5 perfis, denominados de caso;

- Caso 1 - Está cursando atualmente 1 curso de idioma;
- Caso 2 - Está cursando atualmente mais de 1 curso de idioma;
- Caso 3 - Já fez curso de idioma, no entanto não o concluiu e não está cursando atualmente;
- Caso 4 - Já concluiu o curso de idioma, e atualmente não está cursando nenhum outro curso de línguas estrangeiras;

- Caso 5 – Nenhum dos casos anteriores.

Era solicitado que o participante informasse em qual Caso ele se enquadrava, e a partir desta resposta o questionário era direcionava para seções específicas. Essas seções eram semelhantes entre si, uma vez que maioria das perguntas eram análogos, mudando apenas o tempo verbal, entre presente e passado. Também se usava do artificio de apresentar instruções complementares referentes ao seu preenchimento (o que basicamente diferenciava entre o preenchimento do Perfil Caso 1 e Perfil Caso 2), com o intuito de tornar o texto mais claro e objetivo. No entanto, algumas seções apresentaram perguntas distintas, visto que elas não se enquadravam para todos casos, entretanto buscou-se manter uma coerência, entre as diferentes seções.

Entretanto, se o indivíduo informasse que nunca frequentou um curso de idioma (“Caso 5 – Nenhum dos casos anteriores”) o questionário era finalizado, posto que esta classe não será estudada.

Outro aspecto chave dos questionamentos, seria relativo a modalidade do curso, se era presencial ou à distância, entendendo-se que os fatores causadores da evasão se divergem em alguns pontos de acordo com a modalidade, por esta razão, foi preciso abordar questionamentos específicos, sendo assim foi criada uma estrutura de funcionamento correlata ao explicado na situação anterior.

O questionário era composto de duas perguntas com respostas opcionais e 32 a 34 perguntas requeridas, ressaltando que a quantidade de questões dependia da modalidade do curso e do caso que o respondente se enquadrava. Os questionamentos eram múltipla escolha (usando a escala Likert quando conveniente) exceto quando se perguntava a idade ou quando o respondente assinalava a opção “outro”, abrindo um campo para a digitação. O formulário foi dividido em quatro etapas metodologias, onde:

- Seção 1: tinha como propósito conhecer o perfil do respondente, como por exemplo: Idade, sexo, renda familiar, estado civil, entre outros.
- Seção 2: buscava identificar informações referentes ao curso, como modalidade, idioma estudado, duração total do curso, etc.
- Seção 3: investigava-se possíveis dificuldade enfrentadas pelos estudantes de acordo com a modalidade. Para os cursos presenciais foram levantados questionamentos entorno das dificuldades de locomoções e horários. Já para

os cursos à distância o enfoque foi de indicar possíveis problemas de acesso e adaptabilidade com a plataforma de ensino.

- Seção 4: estava relacionada as percepções do ser em relação ao curso, como motivação, interação com a turma/professor, satisfação, entre outras. Deve-se ressaltar que o respondente que não estavam cursando atualmente cursos de idioma eram instruídos para responder essa seção de acordo com a sua percepção de quando realizava o curso.

3.3. Descrição dos modelos de análise

Como mencionado na seção anterior, o Perfil Caso 5 (indivíduos que nunca estudaram cursos de idiomas) não faz parte da área de interesse dessa pesquisa, sendo assim os dados referentes a esse perfil foram deletados do banco de dados. Outra problemática foi a baixa representatividade do número de respostas referentes ao Perfil Caso 2 (indivíduos que estão cursando mais de um curso de idiomas atualmente), por essa razão optou em dividir o estudo em apenas 3 classes:

- Classe 1: Estão cursando atualmente um ou mais curso de idioma;
- Classe 2: Já fizeram curso de idioma, no entanto não conclui e não está cursando atualmente nenhum outro curso de idioma;
- Classe 3: Já concluíram o curso de idioma e atualmente e não está cursando atualmente nenhum outro curso de idioma;

Os dados coletados podem ser usados para prever diversos modelos, desde que sejam definidas as variáveis preditoras e as classes ou variáveis alvo. Neste trabalho, optou-se por avaliar dois modelos de previsão:

- Modelo 1: buscava prever qual classe o respondente pertencia, se o mesmo estava cursando, se havia evadido ou se já tinha concluído o curso;
- Modelo 2: buscava apenas prever se o aluno concluiu ou não o curso, por esta razão foi analisado apenas as 244 respostas pertencente a essas classes.

Os modelos citados acima, foram testados em duas configurações distintas:

- Configuração “A”: Eram utilizadas todas as variáveis para predição, totalizando 29 (vinte e nove);
- Configuração “B”: Eram utilizadas apenas variáveis selecionas de acordo com os conhecimentos pessoais e os adquiridos pela pesquisadora durante a realização do trabalho, assim como, foi utilizada a matriz de correlação para auxiliar a tomada de decisão, sendo escolhida apenas 9 (nove) variáveis.

A configuração “A” tem como característica um maior índice de complexidade, pois usa um número maior de variáveis de entrada, enquanto a configuração “B” incorpora conhecimentos prévios no modelo, buscando propiciar uma base de princípios de forma que minimize a complexidade do modelo. Para a configuração “B” foram utilizados os seguintes atributos: Dificuldade de aprendizado; Modalidade; Duração total; Tempo de estudo; Idade que fez o curso; Escolaridade do pai; Idioma; Situações que ocorreram durante o curso; e Material didático / Facilidade de locomoção.

Deste modo, foram totalizados 28 testes, sendo utilizadas 7 técnicas em dois modelos e duas configurações distintas.

3.4. Implementação de técnicas de aprendizado de máquina

A análise dos dados foi feita por meio de uma plataforma computacional *web* chamada *Jupyter Notebook*, optando por usar o *Python* como linguagem de programação. E para a implementação das técnicas de aprendizado de máquina, foi utilizado a biblioteca *Scikit-learn*, uma vez que essa é amplamente usada para esse âmbito, devido a sua ampla acessibilidade e pela simplicidade e eficiência.

O *Jupyter Notebook* é capaz de ler diversas extensões de arquivo, como csv, txt, json, entre outras. Entretanto foi usada a extensão xlsx para a leitura dos dados, uma vez que o *Google Forms* gerou uma planilha com as respostas obtidas dos questionários nesse formato, e pelo fato que esse tipo de arquivo é propicio para administrar dados.

Para aplicar os modelos de aprendizado de máquina, foi necessário um pré-processamento dos dados, pois de acordo com os “Casos” informados, os questionários eram dirigidos para seções específicas, e consequentemente o *Google*

Forms gerou o seu banco de dados de acordo com cada seção. Sendo assim foi feita uma união dos dados, correlacionado não somente as perguntas iguais, mas também aquelas que possuíam um propósito e uma escala semelhante, tendo como objetivo obter um banco de dados único e conciso.

A métrica utilizada para a avaliação do desempenho das técnicas de aprendizado de máquina foi a acuracidade, dada pela seguinte equação:

$$A = \frac{V_p + V_n}{V_p + F_p + F_n + V_n} \quad (1)$$

Na qual “Vp” representa os verdadeiros positivos, “Vn” os verdadeiros negativos, “Fp” os falsos positivos e “Fn” os falsos negativos. Já a letra “A” simboliza a acuracidade, sendo calculada pela razão de acertos pela quantidade total de valores preditos.

3.4.1. Implementação da técnica K-vizinhos mais próximos

Antes de iniciar a implementação do KNN foi necessário fazer a normalização dos dados, pois havia variáveis que possuía um intervalo relativamente amplo, quando comparado com as demais.

A normalização dos dados ocorreu por meio da biblioteca padrão do *Scikit Learn* denominada *Standard Scaler*, na qual é usada seguinte equação como base:

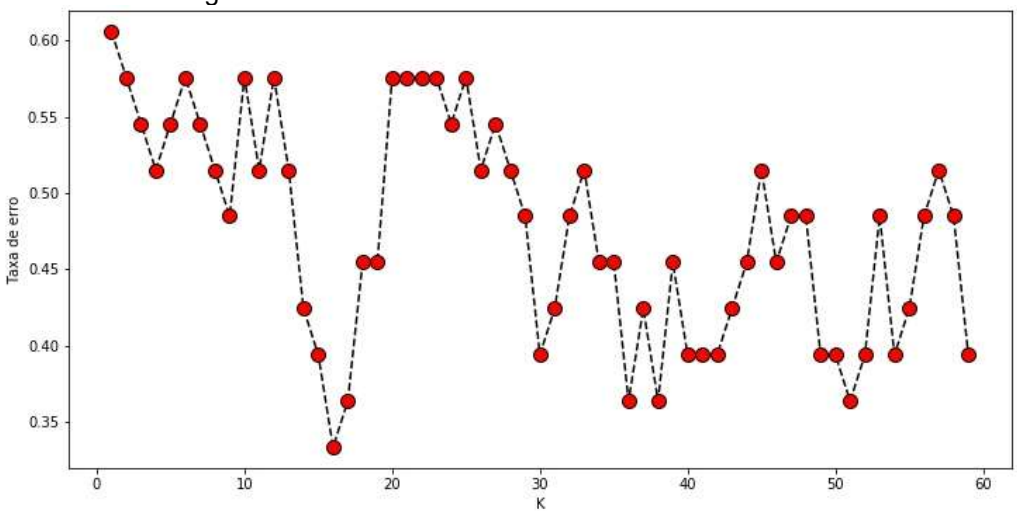
$$\mathrm{z} = \frac{(\mathrm{x} - \mathrm{u})}{s} \quad (2)$$

Onde “x” representa a variável a ser normalizada, “u” é a média e o “s” é o desvio padrão das amostras de treinamento, tendo como intuito de transformar os dados de maneira que a distribuição terá um desvio padrão de 1 e um valor médio de 0.

Outro fator de grande relevância para o estudo é a escolha do K número vizinhos, sendo assim, foi utilizado a plotagem de um gráfico que apresentava as taxas de erros referente ao intervalo de 1 a 60 números vizinhos, sendo escolhido aquele que possui a menor média de erros, buscando dar preferência para os números

ímpares, como meio de evitar empate no momento das classificações. A Figura 14, mostra o exemplo de um dos gráficos plotados.

Figura 14- Taxa de erro versus o K número de vizinhos



Fonte: Elaborada pela autora, 2019

No exemplo ilustrado acima seria escolhido como 17 o número de vizinhos, pois este é o número ímpar que apresenta a menor taxa de erros.

3.4.2. Implementação da técnica de regressão logística

A técnica regressão logística foi aplicada de maneira individual, assim como após um pré-processamento realizado pela técnica Análise de Componentes Principais (PCA). Para a sua execução usou-se a biblioteca *Logistic Regression*, sendo alterado o padrão do *solver* para "*lbfgs*" e do parâmetro *multi_class* de "*warn*" para "*multinomial*", uma vez que esse método permite a generalização da regressão logística para problemas que são multiclasses.

3.4.3. Implementação da técnica de máquina de vetores de suporte

A implementação da técnica de máquina de vetores de suporte se sucedeu através da biblioteca SVC da *Scikit Learn*, no entanto com o objetivo de encontrar melhores resultados, foi definido a Função de Base Radial (RBF, tradução livre do termo inglês *Radial Basis Function*) como o *kernel* para treinar os dados.

Também foram testados os valores do parâmetro “C” que eram alterados de 10^{-1} a 10^{3}, que consiste em modificar as margens da função de decisão, na qual quanto maior o valor do parâmetro, menor será a distâncias entre as margens.

Outro parâmetro alterado foi o “*Gamma*”, na qual os seus valores variavam de 1^{10} a 10^{-4}, definindo assim o nível de influência de um único exemplo de treinamento, quanto maior o valor do parâmetro “*Gamma*” maior será a influência dos pontos que estiverem mais perto dele.

3.4.4. Implementação da técnica árvore de decisão

Para a utilização da técnica na plataforma *Python* foi empregada a biblioteca *Decision Tree Classifier*, sendo aplicado os seus padrões sem alterar nenhum parâmetro. Deste modo foi usada e estratégica de escolher a melhor divisão dos nós, sendo que 2 era o número mínimo de amostras necessárias para a divisão de um nó interno, enquanto para a divisão das folhas era necessário apenas de 1 amostra. Sendo utilizado o índice de pureza Gini para a análise de informação.

3.4.5. Implementação da técnica florestas aleatórias

A implementação da técnica florestas aleatórias foi realizada no mesmo arquivo (“*notebook”)* que a técnica árvore aleatória, por esta razão essas técnicas compartilham a mesma base de dados para treinamento e teste.

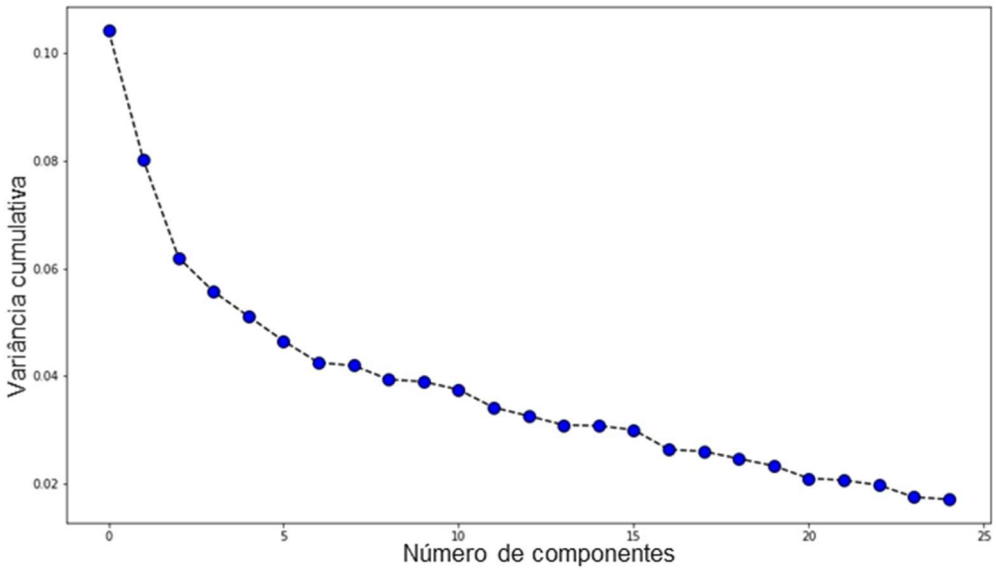
Foi usada a biblioteca *Random Forest Classifier* para a execução da técnica, sendo modificado o parâmetro do número de árvores na floresta para 600, no entanto deve-se ressaltar que quando maior esse número mais pesado computacionalmente será o seu algoritmo, no entanto normalmente tende a melhorar a performance e a estabilidade do mesmo.

3.4.6. Implementação da técnica Análise de componentes principais

Incialmente foi feita a normalização dos dados afim de evitar que variáveis com alta variância tivesse uma influência dominante. Após essa fase, se buscou decidir o número de componentes principais, isso foi feito através da observação do

scree plot que apresentava a porcentagem de variância explicada para cada um dos componentes, conforme mostrado na Figura 15.

Figura 15 - *Scree plot* dos porcentuais de variância explicada

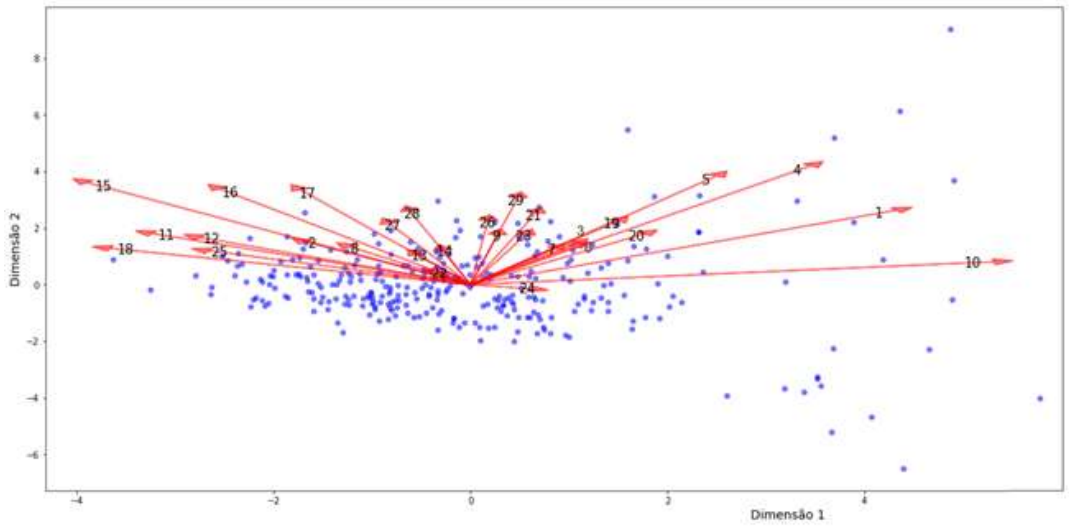


Fonte: Elaborada pela autora, 2019

Para analisar o gráfico acima se usou o método cotovelo que consiste em identificar no gráfico algum ponto em que o ganho marginal diminuíra, dando a origem um ângulo no gráfico. No entanto essa é uma abordagem subjetiva, podendo ter ambiguidade de identificação. Como no exemplo da Figura 15, em que mostra a formação de um ângulo em mais de um ponto.

Também foi realizada uma análise exploratória dos principais através da plotagem do *biplot,* sendo que as amostras foram representadas pelos marcadores azuis, enquanto as variáveis foram representadas por vetores em vermelho (Figura 16).

Figura 16 – *Biplot* de todos os componentes



Legenda	
1 - Caso	16 – Modalidade
2 - Idade	17 - Facilidade com tecnologia/ Distancia
3 - Renda	18 - Material didá¡tico / Facilidade de locomoção
4 - Sexo	19 - Meio de transporte
5 - Estado Civil	20 – Motivação
6 - Filhos	21 – Satisfação
7 - Escolariade	22 - Relevância para a carreira
8 - Escolaridade mãe	23 - Relevância vida pessoal
9 - Escolaridade pai	24 - Dificuldade de aprendizado
10 - Idioma	25 - Nivelamento
11 - Idade que fez o curso	26 - Dificuldade de conciliar a carga horÃ¡ria
12 - Tempo de estudo	27 - Funcionalidades do curso/ Horarios da aula
13 - Duração total	28 - Metodologia/ Integração
14 - Valor da mensalidade	29 - Disciplina/ Relacionamento com o professor
15 - Situações que aconteceram	30 - Satisfação com o IE

Fonte: Elaborada pela autora, 2019

A intepretação do *biplot* possibilita identificar quais os atributos têm um maior efeito em cada dimensão, ou seja, quanto maior o vetor maior será a sua influência, e vice-versa. A Figura 16 mostra que o atributo Idioma (10) possui uma forte influência na dimensão 1. Também pode-se observar a dispersão dos dados, uma vez que um ponto está afastado dos demais, pode ser um indício de um *outlier*.

Após realizadas todos esses estudos foi processada a técnica por meio da biblioteca denominada PCA, sendo definido como 13 o número de componentes principais para os modelos 1A e 2A e 5 para os modelos 1B e 2B. E logo em sequência foi aplicada a técnica Regressão Logística, conforme relatado na seção 3.3.2.

3.4.7. Implementação da técnica Redes neurais

Usou se a biblioteca *MLP Classifier* (*Multi-layer Perceptron classifier* em tradução livre: Classificador *Perceptron* multicamadas), na qual este modelo possui uma capacidade de aprender com modelos não lineares.

Realizou-se uma série de modificações dos parâmetros padrões, como alteração do *solver*, número de neurônios na camada oculta, a forma de ativação da camada oculta, da taxa de aprendizado inicial, parâmetro de penalidade e número máximo de interação, no entanto usou a configuração padrão para os modelos 1B e 2B pois apresentou um melhor desempenho, enquanto para os modelos 1A e 2A alterou o solver de “*adam*” para *'lbfgs'*.

4. RESULTADOS E DISCUSSÕES

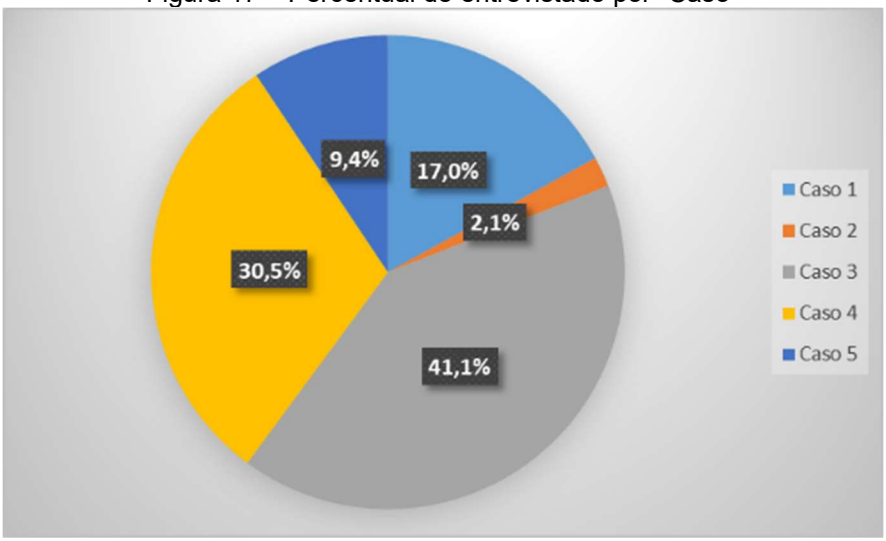
Neste capítulo será apresentado e discutido os resultados obtidos nesse trabalho.

4.1. Caracterização dos participantes da pesquisa

No primeiro momento foi realizado uma análise exploratória dos dados, buscando interpretar e compreender seu comportamento e tendências, afim de extrair informações relevantes para a criação dos modelos de implementação. Para isso foi utilizado principalmente técnicas gráficas, como mapa de calor, *boxplot*, histogramas, e gráfico de pizza.

Como a definição de “Casos” é um dos aspectos chave para o estudo, conforme já mencionado, a análise exploratória partiu desse quesito, sendo plotado o gráfico de pizza, como mostrado na Figura 17:

Figura 17 – Percentual de entrevistado por “Caso”



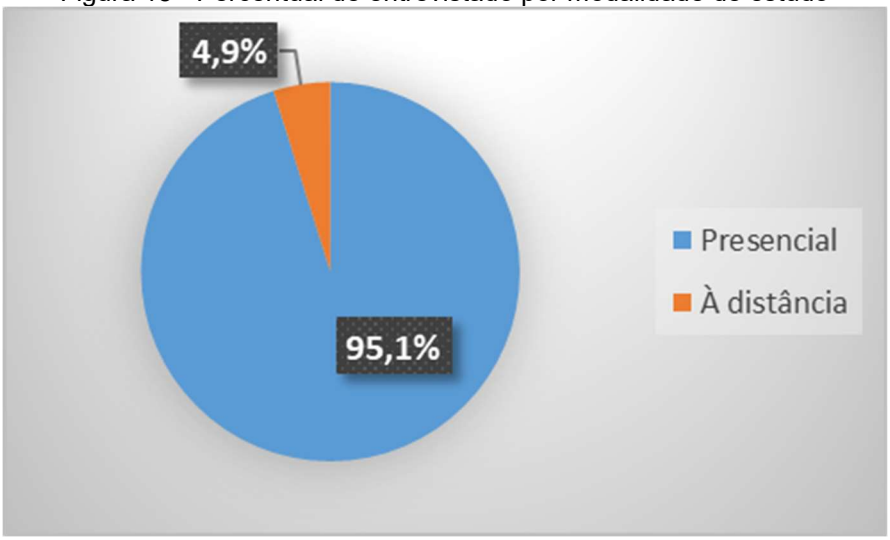
Fonte: Elaborada pela autora, 2019

Ao analisar o gráfico acima é possível verificar que o “Perfil Caso 2” recebeu uma pequena taxa de respondente, o que torna inviável uma análise abrangente desse perfil. Por esta razão, foi decidido que o “Perfil Caso 1” e o “Perfil Caso 2” fossem trabalhados como sendo pertencente a uma mesma classe, uma vez que os mesmos estão atualmente estudando algum idioma. A análise do gráfico também

reforça a premissa que há uma alta taxa de estudantes que não concluem os cursos de línguas estrangeiras, tornando um fator alarmante para a educação brasileira. Ressalta-se que foram excluídos os dados referentes ao “Perfil Caso 5”, sendo assim as próximas análises não englobaram esse perfil.

Também foi analisado a quantidade de respondente de acordo com a modalidade do curso (Figura 18), entendendo que as dificuldades enfrentadas por aqueles que estudam à distância possam a vir ser diferente daqueles que frequentam cursos presenciais.

Figura 18 - Percentual de entrevistado por modalidade de estudo

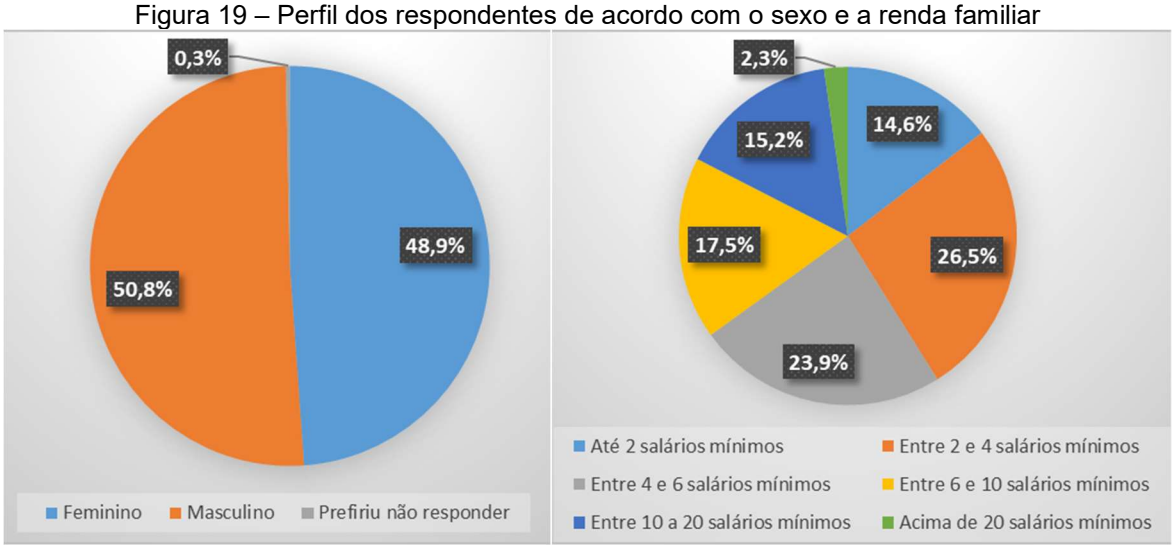


Fonte: Elaborada pela autora, 2019

Verificando os percentuais mostrado na Figura 18, pode-se constatar que o curso à distância ainda apresenta uma baixa representatividade no banco de dados analisado. Por esse motivo preferiu-se fazer uma análise conjunta, sem distinguir os cursos presenciais dos à distância, buscando uma correlação entre as modalidades a fim de tornar a análise mais simples e efetiva.

Com intuito de compreender o perfil dos respondentes, foi analisado principalmente o sexo e a renda familiar (Figura 19), uma vez que estes aspectos foram apontados por autores como sendo aspectos relevantes para evasão escolar.

Figura 19 – Perfil dos respondentes de acordo com o sexo e a renda familiar

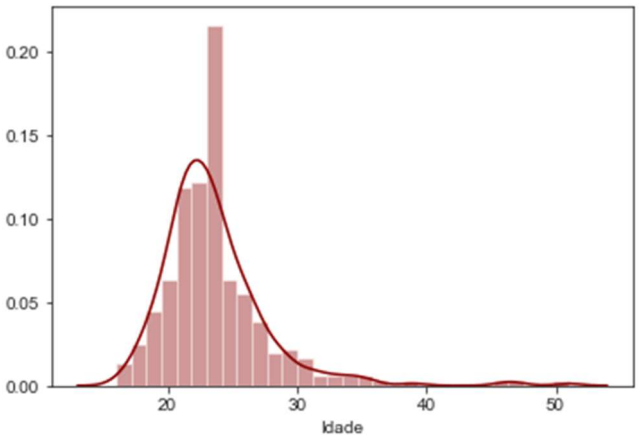


Fonte: Elaborada pela autora, 2019

Quanto ao sexo dos entrevistados, pode-se afirmar os dados estão bem distribuídos, na qual apenas um dos entrevistados, preferiu não se identificar. Enquanto no quesito renda familiar houve um baixo percentual de respondentes que afirmaram que possuem renda acima de 20 salários mínimos, no entanto isso já era esperado, uma vez que apenas 10% da população concentra 43% da soma de rendimentos do país. (SILVEIRA, 2017). Deste modo, acredita-se que esse fator não causará interferência negativa nas análises.

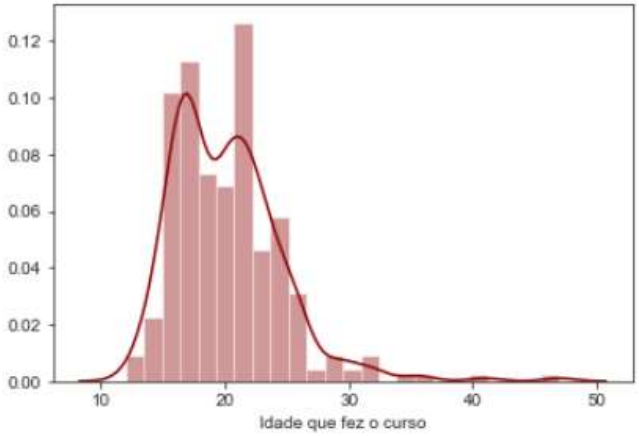
Ciente que os alunos tendem a enfrentar diferentes adversidades dependendo de sua idade, buscou se analisar como estavam dispostas as idades atual e aquela em que o respondente realizou o curso. Para isso, foram feitos os histogramas da Figura 20 e da Figura 21.

Figura 20 - Histograma com a idade atual .



Fonte: Elaborada pela autora, 2019

Figura 21 Histograma com idade do participando de quando foi realizado o curso.

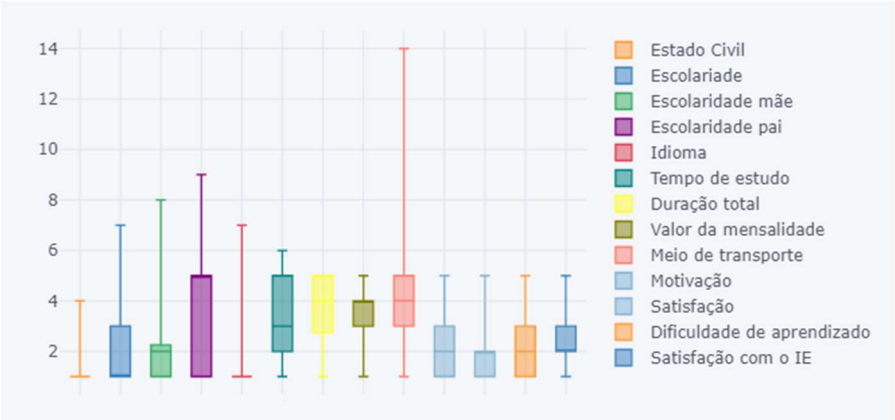


Fonte: Elaborada pela autora, 2019

É possível notar que a grande maioria dos respondentes possuíam idades entre 12 a 28 anos quando realizaram o curso, enquanto a idades atuais variaram entre 18 a 30 anos, isso pode ser atribuído ao fato que a divulgação ocorreu principalmente no meio universitário e pela rede de contatos da pesquisadora. Entretanto, se crê que o número de estudantes de cursos de idiomas está em sua maioria distribuídos nessa faixa etária de idade, uma vez que o aumento pela procura de cursos de línguas muito se deu devido a globalização.

Para detectar de maneira visual as variações sofridas nas amostras, foi plotado um *boxplot* de todas as variáveis, no entanto devido à dimensão optou-se em trazer para o presente trabalho a plotagem de apenas algumas variáveis selecionadas pela pesquisadora, conforme mostrado na Figura 22.

Figura 22 - *Boxplot*

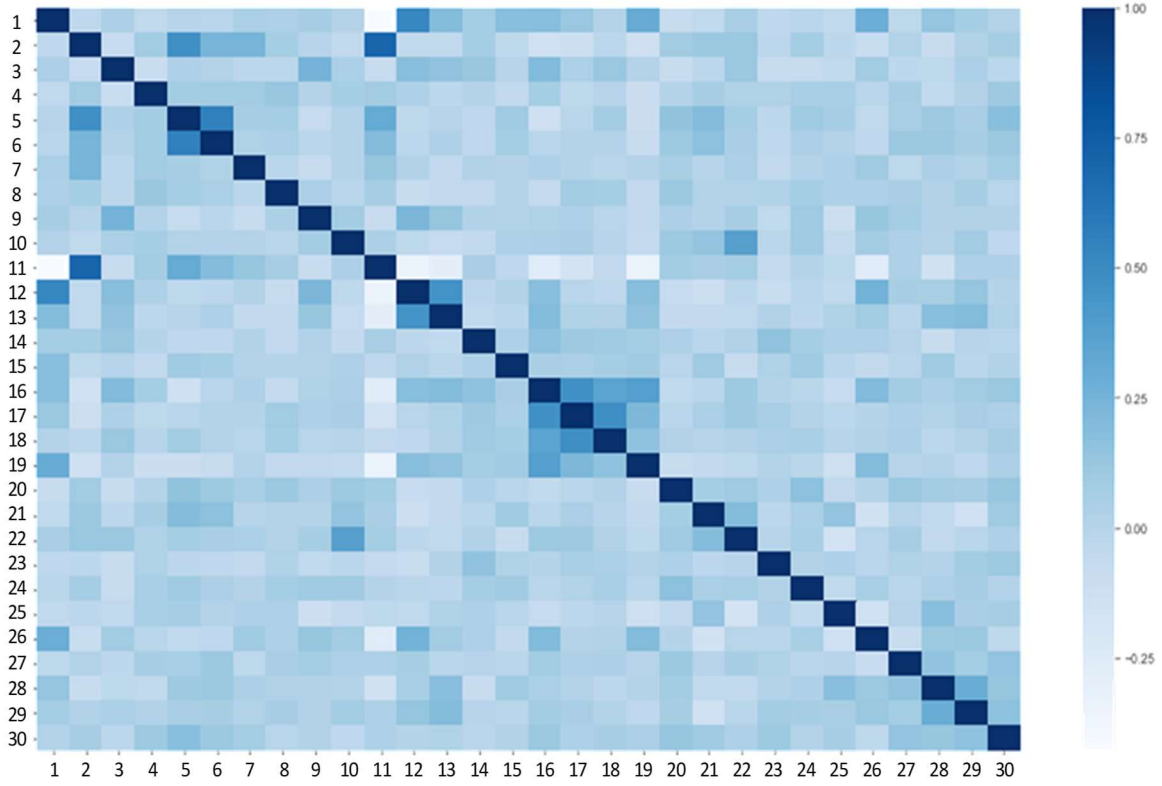


Elaborado pela autora

Ao verificar a Figura 22, nota-se que a variação dos dados é relativamente baixa, uma vez que a maioria dos questionamentos eram de múltipla escolha, contribuindo para essa tendência.

Para identificar as relações entre as variáveis, foi calculada uma matriz de correlação dos dados amostrais e posteriormente foi plotado um mapa de calor divididos por *cluster* (tradução livre: grupo), com o intuito de criar uma representação visual, onde cada célula é exibida de uma cor distinta, sendo essa proporcional à sua posição ao longo de um gradiente de cores, referido na legenda. Enquanto a ordem das linhas se dá através das análises de *cluster* hierárquicos, conforme mostrado na Figura 23.

Figura 23 - Mapa de calor subdivido de acordo com os *clusters*



Legenda	
1 - Caso	16 – Modalidade
2 - Idade	17 - Facilidade com tecnologia/ Distancia
3 - Renda	18 - Material didá¡tico / Facilidade de locomoção
4 - Sexo	19 - Meio de transporte
5 - Estado Civil	20 – Motivação

6 - Filhos	21 – Satisfação
7 - Escolariade	22 - Relevância para a carreira
8 - Escolaridade mãe	23 - Relevância vida pessoal
9 - Escolaridade pai	24 - Dificuldade de aprendizado
10 - Idioma	25 - Nivelamento
11 - Idade que fez o curso	26 - Dificuldade de conciliar a carga horÃ¡ria
12 - Tempo de estudo	27 - Funcionalidades do curso/ Horarios da aula
13 - Duração total	28 - Metodologia/ Integração
14 - Valor da mensalidade	29 - Disciplina/ Relacionamento com o professor
15 - Situações que aconteceram	30 - Satisfação com o IE

Fonte: Elaborada pela autora, 2019

Ao observar o mapa de calor é possível notar que há uma correlação entre a idade atual do indivíduo, estado civil, quantidades de filhos e a idade que foi realizado o curso de idioma. Confirmando a percepção que geralmente as pessoas mais jovens tendem a ser solteiros, e que as pessoas casadas normalmente possuem mais filhos do que aqueles que não possuem relacionamento estável.

Com o propósito de compreender o nível de dificuldade de aprendizado apresentada pelos alunos, foi feita a Tabela 3:

Tabela 3 - Percentual de nível de dificuldade experimentada pelos estudantes de acordo com a classe.

	Classe 2	Classe 3
Muito Baixa	6%	7%
Baixa	26%	36%
Nem baixa, nem alta	34%	47%
Alta	29%	9%
Muito Alta	5%	2%

Fonte: Elaborada pela autora, 2019

A Tabela 3 aponta que apenas 11% dos respondentes que concluíram o curso possuíam alta ou muito alta dificuldade de aprendizado, versus um percentual de 34% quando comparado ao grupo de estudante que evadiram.

Também foi quantificado as idades que os respondentes declaram ter quando concluíram ou não o curso, os resultados são mostrados na Tabela 4:

Tabela 4 - Índices percentuais relativos as faixas etárias de idades que os estudantes possuíam quando realizaram o curso.

Faixa de Idades	Classe 2	Classe 3
de 12 a 17 anos	31%	63%
de 18 a 23 anos	51%	31%
de 24 a 29 anos	15%	7%
mais de 30 anos	3%	0%

Fonte: Elaborada pela autora, 2019

Os resultados acima indicam que o banco de dados apresentou um déficit em relação ao grupo de pessoas que tem idades superior a 30 anos, principalmente quando se diz a respeito daqueles concluíram o curso.

Foi certificado por quanto tempo os alunos estudaram em um curso de idioma, até finalizarem ou evadirem do curso, os resultados são mostrados na Tabela 5.

Tabela 5 - Percentual de respondentes referente ao tempo de estudo.

Tempo de estudo	Classe 2	Classe 3
Até 1 ano	52,9%	4,8%
de 1 a 2 anos	26,4%	16,3%
de 2 a 3 anos	11,4%	6,7%
mais de 3 anos	9,3%	72,1%

Fonte: Elaborada pela autora, 2019

Por meio da Tabela 5 é possível notar que os alunos que estudaram por um maior período de tempo tendem a concluir o curso, dado que 72% dos que finalizaram o curso estudaram por período superior a 3 anos.

Outro aspecto verificado foi quanto ao nível escolaridade apresentados pelos genitores dos estudantes que evadiram e daqueles que concluíram o curso (Tabela 6).

Tabela 6 - Percentual quanto ao nível de escolaridade dos genitores dos respondentes

Nível de escolaridade	Paterna			Materna		
	Classe 1	Classe 2	Classe 3	Classe 1	Classe 2	Classe 3
Não sei informar	2%	2%	4%	0%	0%	0%
Sem escolaridade	0%	1%	0%	0%	1%	0%
Ensino Fundamental Incompleto	5%	15%	6%	2%	8%	5%

Ensino Fundamental Completo	5%	4%	4%	5%	6%	1%
Ensino Médio Incompleto	6%	8%	2%	2%	2%	4%
Ensino Médio Completo	28%	29%	27%	18%	37%	25%
Ensino Superior Incompleto	6%	6%	11%	9%	4%	7%
Ensino Superior Completo	48%	30%	40%	58%	37%	50%
Mestrado ou Doutorado	2%	5%	7%	6%	4%	9%

Fonte: Elaborada pela autora, 2019

É possível notar que os maiores percentuais de cada classe estão localizados majoritariamente nos níveis de Ensino Médio e Superior completo.

Foi verificado a relação entre as classes estudadas com o idioma em estudo, conforme mostrado na Tabela 7:

Tabela 7 - Percentual entre o Idioma estudo e classe dos respondentes.

	Classe 1	Classe 2	Classe 3
Alemão	3	7	4
Espanhol	2	4	6
Francês	5	13	1
Inglês	52	115	90
Italiano	3	0	2
Japonês	0	1	0
Mandarim	0	0	1

Fonte: Elaborada pela autora, 2019

Conforme pode ser visto nos dados da Tabela 7, a grande maioria dos respondentes estudaram o idioma inglês, uma vez que este é considerado como sendo uma língua utilizada em diferentes partes do mundo.

Também foi levantado as situações que ocorreram durante a realização do curso, no entanto como esse havia um campo livre para digitação, se recebeu os mais diversos relatos, deste modo foram selecionadas apenas algumas variáveis para compor a Tabela 8. Deve-se ressaltar que a mesma pessoa tinha a liberdade de apontar mais de uma situação.

Tabela 8 - Situações que ocorreram durante a realização do curso de idioma

Situações	Classe 2	Classe 3
Iniciou outros cursos (de idiomas ou não)	36	25
Começou a trabalhar	33	13

Mudou-se de endereço	27	16
Enfrentou problemas de saúde	13	8
Realizou viagens	10	32
Falta de tempo	5	0
Casou-se/Iniciou união estável	3	0
Dificuldade em conciliar o estudo de idiomas com a universidade	3	0
Curso era muito básico	1	0
Falta de condições para pagar o curso	2	0
O curso não cumpriu com o combinado	1	0
Separou-se/divorciou-se	1	0

Fonte: Elaborada pela autora, 2019

Ao observar a Tabela 8 é possível notar que a classe 2 foi que apresentou mais situações, uma vez que alguns respondentes utilizaram esse campo para justificar o motivo de sua evasão.

E por fim foi analisado quanto as dificuldades apresentadas pelos cursos presenciais e a distância, unindo os questionamentos em torno da facilidade de locomoção, com o questionamento quanto o grau de dificuldade com os materiais didáticos e comunicação com o tutor. Para isso foi necessário transformar as respostas em uma mesma escala, logo o resultado pode ser visto na Tabela 9.

Tabela 9 – Dificuldades com locomoção, matérias didáticos e comunicação com o tutor.

	Classe 2	Classe 3
Muito fácil	15%	13%
Fácil	46%	56%
Nem difícil, nem fácil	24%	30%
Difícil	12%	1%
Muito Difícil	2%	0%

Fonte: Elaborada pela autora, 2019

É possível observar que os resultados mostram que nenhum dos entrevistados pertencente a classe 3 apresentou muita dificuldade nas modalidades analisadas.

4.2. Comparação e análise dos classificadores

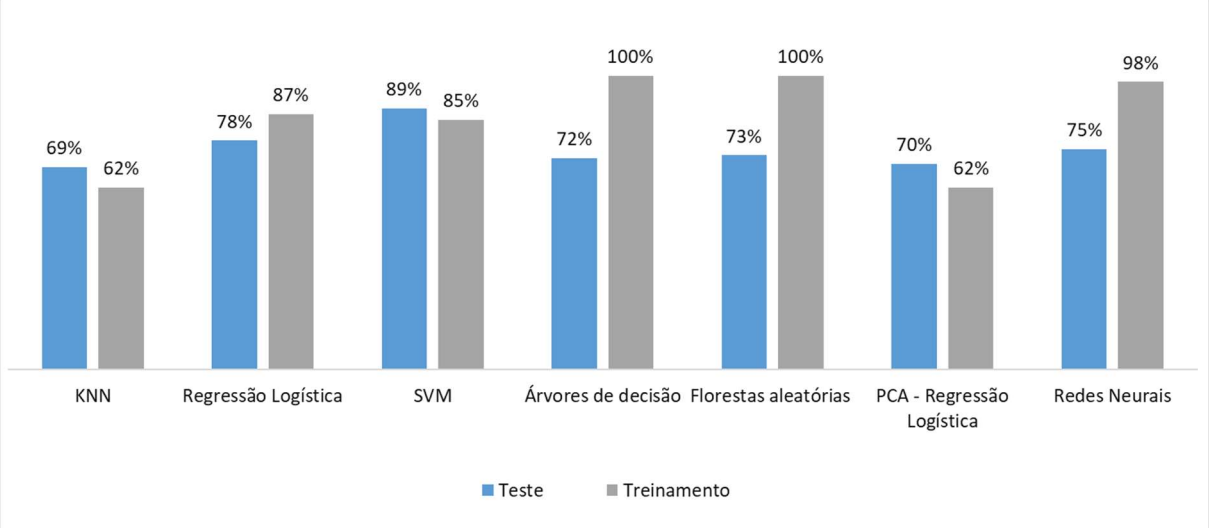
Após realizada a implementação das técnicas de aprendizado de máquina, os resultados tanto da fase de teste como a de treinamento foram analisados utilizando a métrica de acuracidade descrita pela Equação (1). As avaliações foram realizadas

de acordo como os modelos de análises e as suas respectivas configurações conforme referidas na seção 3.4.

4.2.1. Resultados do Modelo 1A

Foi iniciada as análises pelo Modelo 1 configuração A, que foi considerado como sendo o modelo mais abrangente dentre os analisados, uma vez que utilizada maior número de variáveis para a predição e pelo fato de buscar prever todas as três classes, salientando que isso gera um modelo mais complexo e computacionalmente mais lento. Entretanto, como foi trabalhado com um banco de dados relativamente pequeno, essa problemática apresentou baixa relevância para o estudo atual. A Figura 24 mostra os percentuais de acuracidade desse modelo, tanto para os dados de teste como para os dados de treinamento.

Figura 24 – Percentuais de acuracidade das técnicas referente ao modelo 1A



Fonte: Elaborada pela autora, 2019

Conforme mostrado na Figura 24, as técnicas Máquina de Vetores de Suporte e Regressão Logística foram as que apresentaram melhor desempenho no Modelo 1A, obtendo uma acuracidade máxima de 89% e 78%, consecutivamente. Enquanto a técnica K-vizinhos Mais Próximos (com 37 vizinhos) foi o que obteve a pior performance.

Também foi plotado as precisões das técnicas para os dados de treinamento, tendo como objetivo verificar a diferença de percentuais entre a fase de treinamento

e a de teste, uma vez que por meio dessa análise pode-se verificar se as técnicas possuem indícios de sobre-ajuste ou sub-ajuste.

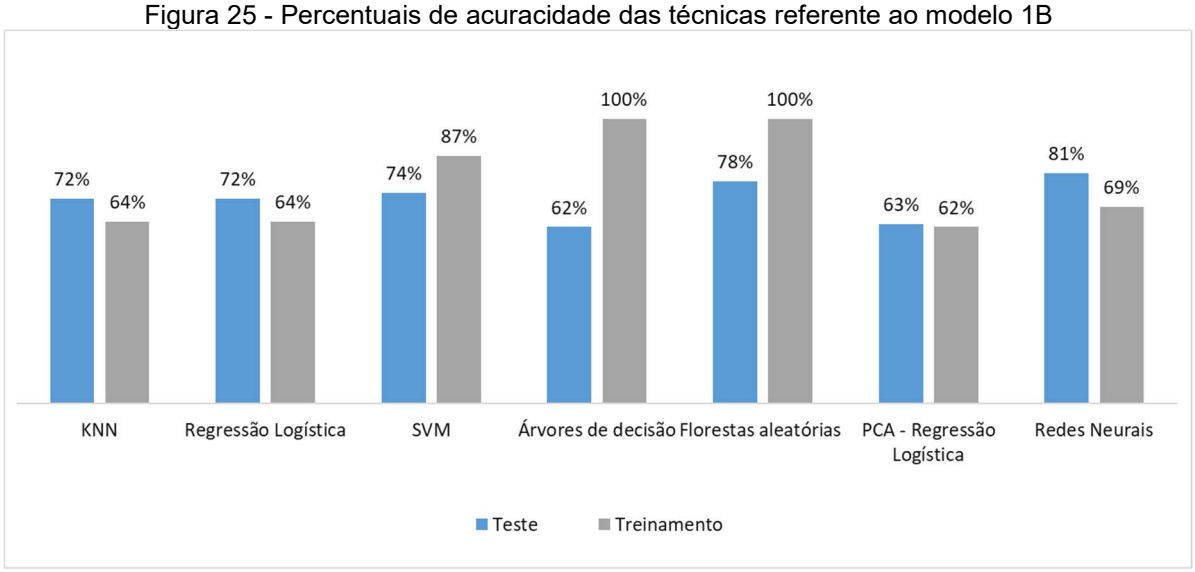
Estudando os resultados acima, as técnicas Árvore de Decisão e Florestas Aleatórias foram o que apresentaram maiores diferenças percentuais entre de previsão dos dados de teste e treino, que pode ser atribuído ao fato de que não foi utilizado nenhum mecanismo de “poda”, ou técnica que buscasse evitar esse acontecimento. Os resultados também demostram que o SVM possuiu uma ótima capacidade de generalização, uma vez que a acuracidade da fase de teste foi maior que a obtida na fase de treinamento.

Como indicador geral, foi calculado a média aritmética simples de acuracidade do modelo por meio da soma dos percentuais de acuracidade obtidos dividindo pela quantidade de técnicas, que no caso eram 7, obtendo aproximadamente 75% de média para o Modelo 1A.

4.2.2. Resultados do Modelo 1B

O modelo 1B por sua vez também buscou prever as 3 classes, contudo com o objetivo de minimizar a complexidade do modelo e analisar os atributos de maior impacto para as previsões, foi selecionado 9 atributos, baseando-se nos conhecimentos adquiridos pela pesquisa, assim como pela análise exploratória dos dados. Os resultados desse método podem ser observados na Figura 25:

Figura 25 - Percentuais de acuracidade das técnicas referente ao modelo 1B



Fonte: Elaborada pela autora, 2019

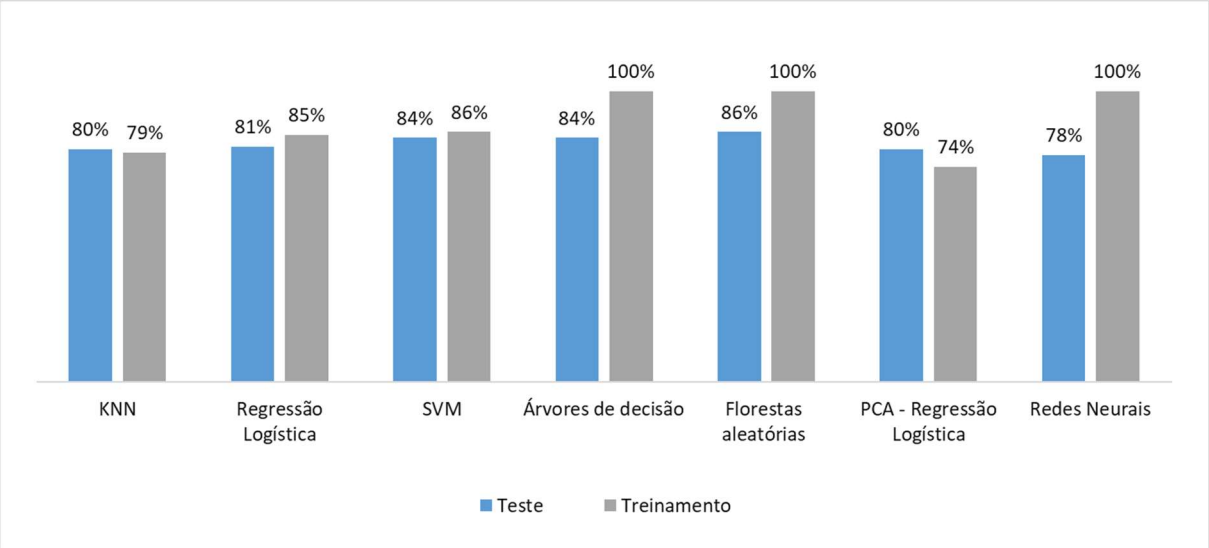
Ao analisar os resultados demostrado na Figura 25 mostra que o modelo ilustrado apresentou uma acuracidade máxima de 81%, alcançada por meio das Redes Neurais e uma acuracidade mínima de 62% tida pela técnica Árvores de Decisão, resultando em um média geral de 72% de acuracidade. Quando comparamos esses índices com os resultados apresentados pelo Modelo 1A, notamos que este modelo de maneira geral apresentou resultados inferiores, com exceção das técnicas K-vizinhos Mais Próximos, Florestas Aleatórias e Redes Neurais. Entretanto deve-se frisar que foi utilizado apenas 31% dos atributos de entrada para gerar a previsão da configuração B.

4.2.3. Resultados do Modelo 2A

Diferente do primeiro modelo, o modelo 2 buscou prever apenas se o estudante havia ou não concluído o seu último curso de idioma em sua totalidade, sendo assim utilizou os dados apenas destas duas classes.

Acreditasse-se que este modelo seja um método mais prático para ser utilizado como forma de previsão no dia-dia. Logo, o modelo também foi executado nas duas configurações, sendo que os resultados obtidos pela configuração A está apresentado na Figura 26:

Figura 26 - Percentuais de acuracidade das técnicas referente ao modelo 2A



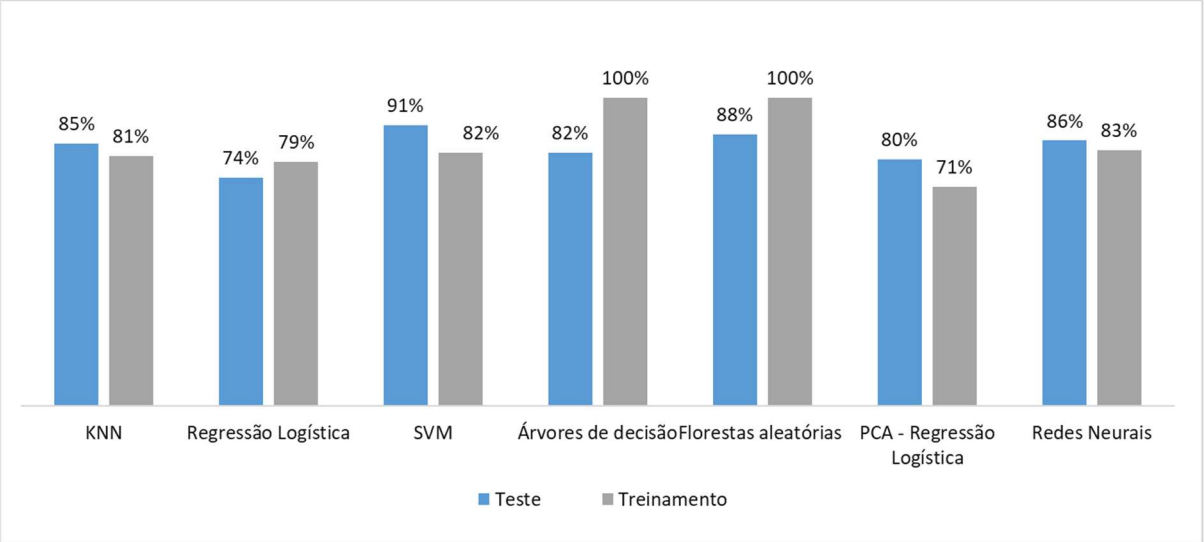
Fonte: Elaborada pela autora, 2019

Verificando a Figura 26 é possível observar que houve baixa variação entre os percentuais de acuracidade dos dados de treinamento com os de teste, assim como entre as técnicas implementadas. Esse modelo obteve uma acuracidade média de 82%, em que a técnica Florestas Aleatórias apresentou uma acuracidade de 86%, sendo a técnica que demostrou a melhor performance do modelo, seguida pela técnica Árvore de Decisão com 84%, na qual ambas as técnicas compartilhavam a mesma base de dados de treino e teste, o que não acontecia para as demais, uma vez que as outras técnicas foram implementadas em diferentes arquivos, onde a divisão dos dados para treinamento e teste eram realizadas de maneira randômica.

4.2.4. Resultados do Modelo 2B

E por fim foi implementado o modelo 2B, que conta com o menor nível de complexidade computacional quando comparado com os demais modelos, visto que possuí um menor número de variáveis de entrada quando comparado com a configuração “A”, assim como apresenta uma quantidade menor de classes que devem ser previstas quando comparado com o Modelo 1. Os resultados atingidos por esse modelo estão dispostos na figura 27.

Figura 27 - Percentuais de acuracidade das técnicas referente ao modelo 2B



Fonte: Elaborada pela autora, 2019

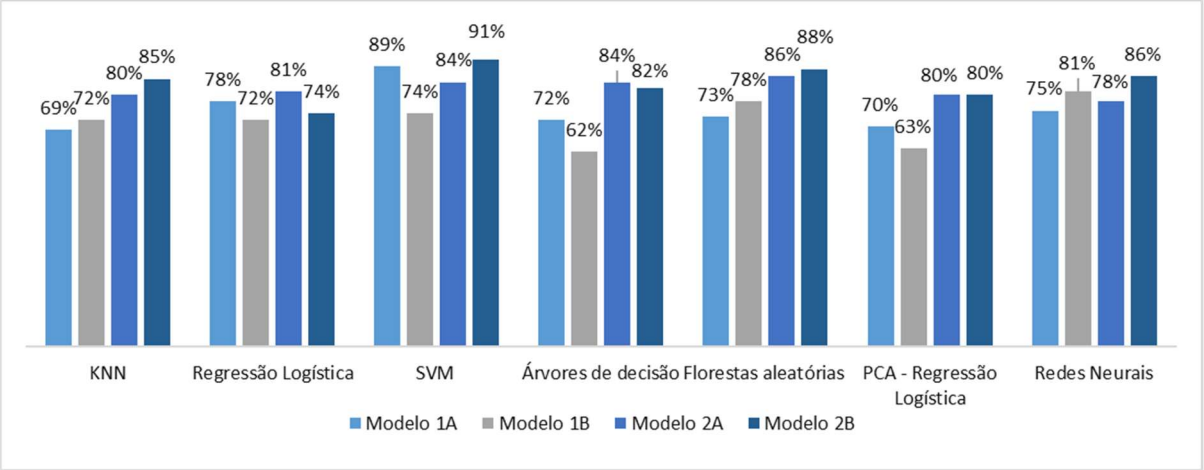
O modelo 2B foi o que apresentou a maior média aritmética entre os modelos, tendo um percentual de 84% de acuracidade, atingindo uma acuracidade máxima de 91% de acertos por meio da técnica Máquina de Vetores de Suporte. Outro aspecto

que chama atenção é o fato que ao usar a técnica Análise de Componentes Principais conseguiu melhor o desempenho apresentado pela técnica Regressão Logística mesmo parametrizado para selecionar apenas 5 atributos de entrada

4.2.5. Comparativos dos resultados

Com o propósito de fazer um comparativo e um estudo das eficiências das técnicas de aprendizado de máquina foi plotada a Figura 28 que mostra o comportamento da acuracidade das mesmas de acordo com os modelos.

Figura 28 - Comparativo dos percentuais de acuracidade obtidos pelas técnicas de aprendizado de máquina



Fonte: Elaborada pela autora, 2019

Observando o comportamento do gráfico pode-se observar que de maneira geral as técnicas apresentam melhor resultado de acordo com as configurações, ou seja há técnicas que trabalharam melhor com um conjunto maior de dados, como teve aqueles que se adaptaram melhor com uma quantidade reduzida de atributos. Também é possível verificar que a configuração “A” apresentou melhor desempenho para o Modelo 1, enquanto a configuração “B” obteve melhor desempenho para o modelo 2. Ainda quando comparado os modelos nota-se que o Modelo 2 foi o que apresentaram maior média de desempenho, como já era esperado pois há uma classe a menos para ser predita.

A técnica Máquina de Vetores de Suporte implementado no modelo 2B foi o que apresentou a maior índice de assertividade de 91%. Por esta razão optou em

realizar uma análise mais aprofundada do modelo, sendo plotado uma matriz de confusão (Figura 29) com os valores reais versus valores predito na fase de teste, onde sua diagonal principal indica a quantidade de acerto, e a soma dos demais algarismos da matriz são a quantidade de erros.

Figura 29 - Matriz de confusão dos valores reais versus os preditos na fase de teste

Fonte: Elaborada pela autora, 2019

Observando a matriz acima é possível verificar que a classe 3 foi a que apresentou a maior taxa de erro, na qual dos 39 estudantes desistentes do curso, 4 destes foram classificados como concluintes, resultando em um percentual de erro de aproximadamente 10%. Contudo, ressalta-se que não ocorreu grandes variações percentuais quanto as taxas de erros entre as duas classes.

4.2.6. Resultados das técnicas de aprendizado não supervisionado

A fim de curiosidade foram usadas as técnica Máquina de Vetores de Suporte e Florestas Aleatórias nos Modelos 2A e 2B (Tabela 10) para prever se os alunos que estão cursando irão evadir ou se vão concluir o curso. Para isso foi utilizada a base de dados completa das classes 2 e classe 3 para treinamento do modelo, e a classe 1 foi usada como teste.

Tabela 10 - Resultados referentes aos Modelo 2A e 2B

Modelos	Técnicas	Classe 2	Classe 3	Acuracidade do Treinamento

2A	SVM	57	8	91%
	Florestas aleatórias	58	7	100%
2B	SVM	57	8	84%
	Florestas aleatórias	59	6	100%

Fonte: Elaborada pela autora, 2019

A Tabela 10 aponta previsões semelhantes entre si, indicando quase 90% dos alunos que responderam o questionário tendem a não concluir o curso. Também é possível observar que a técnica Máquina de Suporte de Vetores não teve variação entre as previsões obtidos nas diferentes configurações. Deve ressaltar que a acuracidade indicada na Tabela 10 refere-se ao treinamento, uma vez que não é possível verificar a dos dados de teste, por se tratar de um aprendizado não supervisionado.

5. IMPLICAÇÕES PRÁTICAS

A configuração B apresentou bons resultado utilizado cerca de apenas 31% dos atributos de entrada para gerar a previsão, por esta razão acredita-se que as variáveis selecionadas representem bem o conjunto de dados, tendo essas maiores relevâncias para o estudo de evasão. Portanto foi elaborada a Tabela 11 contendo sugestões de como as instituições podem proceder, frente a essas problemáticas.

Tabela 11 - Sugestões de medidas que podem ser tomadas pelas instituições de ensino de idiomas

Problemáticas	**Sugestões**
Dificuldade de Aprendizado	É importante que os professores acompanhem a trajetória dos seus alunos, oferecendo um apoio adicional e estimulo, mostrando que eles são capazes de ampliar o seu conhecimento, desenvolver e ampliar o seu potencial.
Modalidade de Ingresso	Os cursos à distância foram os que apresentaram maior taxa de evasão, sendo assim é necessário que os estudantes recebam incentivos para desenvolver a sua individualidade, autonomia e motivação, uma vez que é necessário o engajamento de ambas as partes para a construção do processo de aprendizado. Outra sugestão seria que os tutores estimulem a interação coletiva, seja ela por meios de comunicação, ou até mesmo presencial, evitando o sentimento de isolamento.
Idade em que se realizou o curso	É essencial que as instituições entendam as problemáticas de cada faixa etária, a fim de criar estratégias direcionada ao seu público, como por exemplo criar diferentes segmentações de metodologia de ensino. Visto que os dados apresentaram que a maior taxa de evasão ocorreu no público com idades mais altas, recomenda-se que as instituições possuam horários flexíveis, e com possibilidade de monitorias presenciais, pois comumente esses indivíduos possuem pouco tempo disponível, e apresentam dificuldades para estudarem com materias complementares disponíveis em meios tecnológicos.
Tempo de Estudo	As instituições devem assegurarem que os alunos criem apenas expectativas coerentes a realidade, de modo que essas possam ser atendidas, pois caso contrário o aluno poderá se sentir desapontado, e posteriormente venha a evadir logo no primeiro ano. Para isso é necessário que os cursos disponibilize informações legitimas e de fácil entendimento e acesso, visando a conscientização de todos.
Duração Total	Os dados coletados apontaram que a maior taxa de evasão ocorreu nos cursos que tinham duração de até um ano, diante desse cenário aconselha-se que as escolas de idiomas, criem políticas de conscientização dos alunos, pois muitos desse procuram esses cursos com a expectativa de aprender um idioma rapidamente e com um falso sentimento que será mais fácil o aprendizado do que quando comparado com os cursos de longa duração. No entanto, isso não é uma verdade, uma vez que é preciso assimilar todo conteúdo em um espaço de tempo menor, a ainda deve-se ressaltar que tempo de aprendizado é dependente do nível de dificuldade experimentado pelo aluno, assim como a sua disponibilidade para se dedicar aos estudos, deste modo, a duração do curso pode se estender, e o aluno necessita estar ciente deste possível evento.

Escolaridade do Pai	As instituições podem desenvolver campanhas tem tenha objetivo de trazer as presenças das figuras tanto materna como a paterna dos estudantes mais jovens, com o intuito de conscientizar não somente os alunos, mas também os seus genitores sobre a importância da educação e do aprendizado de idiomas em si. Já para os alunos com idades mais altas, essa conscientização pode ser feita individualmente, e caso estes tenham filhos e/ou companheiros, o ideal seria demonstrar a relevância e até mesmo incentiva-los a buscarem aprendizado de uma nova língua.
Idioma	Visto que os alunos podem enfrentar diferentes dificuldades referentes ao idioma estudado, sugere-se que as abordagens de ensino sejam distintas, entendendo-se que os alunos tendem a possuírem maior grau de dificuldade em idiomas que se há baixa vivência ou que as origens sejam divertes da língua portuguesa. Nos idiomas que o aprendizado é mais complexo, sugere-se uma maior quantidade de aulas e materias complementares disponíveis.
Situações que ocorreram durante a realização do curso	Os resultados dos questionários apontaram as mais diversas situações que ocorreram durante a realização do curso, sendo assim recomenda-se que as instituições estabeleçam relações próximas aos seus alunos tendo como intuito de resolverem as problemáticas juntos. Como por exemplo, houve alunos que relataram dificuldades com o pagamento das mensalidades, uma das soluções seria aumentar o número de parcelas e consequentemente diminuir o valor mensal, visto a real necessidade do estudante, outro exemplo seria que a instituição poderia reduzir a carga horária semanal e estender a duração total do curso, para aqueles que estão com dificuldades de conciliar os estudos com o seu tempo disponível, é claro que estas sugestões podem gerar ajustes nos valores finais, no entanto isso deverá ser acordado entre as partes interessadas.
Material Didático/ Facilidade de Locomoção	Sabendo-se da relevância dada a facilidade de transporte, recomenda-se que as escolas de idiomas presenciais busquem lugares estratégicos, que possuam fácil acesso, e de preferência que sejam providas por transporte públicos. Já para as instituições de ensino a distância recomenda-se possuam bons material didático e que seja de fácil acesso. Também se aconselha que os canais de comunicação entre estudantes e instituição sejam eficientes, permitindo a troca rápida de informação, assim como este devem proporcionar um ambiente que seja favorável para um bom relacionamento entre as partes.

Fonte: Elaborada pela autora, 2019

Notou-se que a dificuldade de aprendizado enfrentado pelo grupo de respondentes que desistiram do curso foi maior quando comparado com aqueles que concluíram os estudos (Tabela 3). Por esta razão, acredita-se que as instituições de ensino deveriam oferecer um suporte extra a estes alunos, buscando minimizar esses obstáculos.

Já no quesito modalidade, verificou-se que a taxa de evasão é maior para os cursos à distância, uma vez que dos 104 respondentes que declaram ter finalizado o curso, apenas 1 desses estudava à distância, representando menos de 1%, entretanto ressalta-se que o presente estudo contou com um banco de dado reduzido nesse quesito, o que demonstra que os estudantes ainda possuem resistências e dificuldades de adaptação com esse tipo de metodologia.

Também foi possível verificar que mais da metade dos alunos que concluíram o curso de línguas estrangeiras tinham idade de até 17 anos (Tabela 4), caracterizando-se um público mais jovem do que quando comparado com os que desistiram, indicando uma relação entre a idade e a tendência de evadir, uma vez que quanto mais alta a idade da pessoa, menor é o seu tempo disponível para a realização dos cursos.

Outra constatação observada, foi que os estudantes tendem a evadirem dos cursos de idiomas logo no seu primeiro ano de estudo (Tabela 5), o que pode ser explicado pelo fato que estes tendem a serem mais empolgados na fase inicial, entretanto ao decorrer do tempo esse entusiasmo sede lugar para possíveis frustações.

Apurou-se que dentre os 90 respondentes que optaram por cursos de duração superior a 4 anos, deste 56 afirmaram ter finalizado o curso, correspondendo um percentual de 62%, enquanto os cursos com duração de até 1 ano tiveram apenas 33% de taxa de conclusão, podendo ser um indicio que os cursos de longa duração recebam maior credibilidade frente aos alunos.

Além disso, averiguou que cerca de 85% dos alunos que concluíram o curso de idiomas a escolaridade paterna era igual ou superior ao ensino médio completo (Tabela 6), demostrando que os pais possuem um caráter fundamental no incentivo a educação, assim como aponta que os filhos se espelham nos exemplos compartilhados em seu lar.

Apesar do pequeno banco de dados, foi possível averiguar que os idiomas que possuem um certo nível de inteligibilidade com a língua portuguesa a taxa de conclusão tende a ser maior, como no caso do idioma Espanhol e do Italiano (Tabela 7).

Verificou-se por meio da Tabela 8, que todos os alunos que se casaram ou iniciaram uma união estável durante a realização do curso de línguas estrangeiras desistiram do curso, o que pode ser explicado pelo fato que estas pessoas normalmente passam a se dedicar mais ao lar e ao cônjuge, modificando a sua rotina habitual. Também é possível observar que muitos desses alunos que desistiam informaram que iniciaram novos cursos, que começaram a trabalhar, indicando uma possível limitação de tempo disponível para a realização do curso.

E por fim foi analisado quanto ao grau de dificuldade em relação a locomoção para aqueles que realizaram cursos presenciais, e a dificuldade com materias didáticos e comunicação com o tutor para aqueles que estudaram a distância, podendo verificar que apenas 1% dos alunos que concluíram o curso relataram dificuldade quanto aos quesitos avaliados, enquanto os que evadiram possuem um percentual de 14% (Tabela 9). O que indica que esses quesitos podem ser tidos como um dos motivos relevantes para a falta de engajamento escolar.

6. CONSIDERAÇÕES FINAIS

As técnicas de aprendizado de máquina propostas buscaram classificar o status do respondente em relação ao curso de idioma, apresentando duas vertentes, na qual a primeira delas buscava prever se o estudante estava cursando, desistido ou se já havia concluído o curso, enquanto a segunda vertente visava analisar apenas se o aluno havia evadido ou não.

Entre as técnicas apresentadas a Máquina de Vetores de Suporte foi o que atingiu o maior índice de acuracidade de previsão (91%), demostrando uma boa capacidade de generalização. Sendo superada a expectativa, devido à complexidade da problemática em estudo, visto a série de fatores inter-relacionados.

Por meio do aprendizado de máquina não supervisionado visou prever se os alunos que informaram que estão cursando idiomas atualmente irão ou não concluir o curso, obtendo como previsão que quase 90% desses irão evadir, índice que gera preocupação, visto a série de malefícios causados por este acontecimento. Ressaltando assim a relevância do estudo, uma vez que a previsão possa a vir auxiliar no combate da evasão, propiciando a tomada de atitudes proativas.

Acredita-se que a utilização das técnicas de aprendizado de máquina sejam uteis para encontrar boas previsões para a problemática em estudo, bem como presumimos que essas técnicas também possam ser aplicadas em problemas de evasão em outros contextos, como no ensino superior e médio, uma vez que os aspectos apontados como causadores deste evento muito se assemelham entre si.

Sugere-se que em pesquisa futuras sejam utilizadas outras técnicas de Aprendizado de Máquina para a predição da evasão de cursos de línguas estrangeiras assim como na evasão de cursos de ensino médio e superior. Também se sugere a utilização de métodos como Ganho de Informação (tradução livre do termo *Information Gain*) para a formação das Arvores de Decisão, Análise de Correspondência Múltipla (MCA) e Análise Multivariada para a análise de dados.

Ainda, se aconselha que seja realizado um estudo em que o resultado da previsão de evasão seja informado antecipadamente ao professor, a fim de verificar se a informação prévia contribui de fato para a redução das taxas de evasão dos cursos.

REFERÊNCIAS

ARAÚJO, Jaíne Gonçalves. **Evasão Na Ead:** Um Survey Com Estudantes do Curso de Licenciatura em Música a Distância da Unb. Brasília, 2015. Originalmente apresentada como dissertação do Programa de Pós-graduação em Música do Departamento de Música da Universidade de Brasília, 2015.

BARLEM JGT; LUNARDI VL; BORDIGNON SS; BARLEM ELD; LUNARDI FILHO WD; SILVEIRA RS; ZACARIAS CC. Opção e evasão de um curso de graduação em enfermagem: percepção de estudantes evadidos. **Revista Gaúcha Enfermagem**, Porto Alegre (RS) 2012 jun;33(2):132-138.

BAYMA-FREIRE, H.; ROAZZI, A.; ROAZZI, MM. Ou o nível escolar de dois países interfere na permanência dois filhos na escola? || O nível de escolaridade dos pais interfere na permanência das crianças na escola?. **Revista de Estudos e Pesquisas em Psicologia e Educação**, [Sl], v. 2, n. 1, p. 35-40, jul. 2015. ISSN 2386-7418.

BAZZOTTI, C.; GARCIA, E. A Importância do sistema de informação gerencial na gestão empresarial para tomada de decisões. **Ciências Sociais Aplicadas em Revista**, v. 6, n. 11, 2006.

BEZERRA, Maria do Carmo Lima. **Dificuldades de Aprendizagem e os Fatores que Influenciam o Fracasso Escolar**, Itaporanga, Paraíba, 2014. Originalmente apresentada como monografia de especialização para a Universidade Estadual da Paraíba, 2014.

BIAZUS, Cleber Augusto. **Sistema de fatores que influenciam o aluno a evadir-se dos cursos de graduação na UFSM e na UFSC:** um estudo no curso de ciências contábeis. 2004. 203 f. Tese (Doutorado em Engenharia da Produção) - Universidade Federal de Santa Catarina, Florianópolis, 2004.

BORJA, I. M. F. S.; MARTISN, A. M. O., 2014. Evasão escolar: desigualdade e exclusão social. **Revista Liberato**, Novo Hamburgo, v. 15, n. 23, p. 01-104, jan./jun. 2014.

BRAGA, Antônio de Pádua; LUDEMIR, Teresa Bernada, CARVALHO, André Carlos Ponce de Leon Ferreira. **Redes neurais artificiais: teoria e aplicações**. Rio de Janeiro: LTC. 2007.

BREIMAN, L. Random Forests. **Machine Learning**, v. 45, p. 5-32, Outubro 2001. ISSN 1573-0565.

BRITISH COUNCIL.: **Demandas de Aprendizagem de Inglês no Brasil:** elaborado com exclusividade para o British Council pelo Instituto de Pesquisa Data Popular, 1ª Edição ,São Paulo, SP: British Council Brasil, 2014. Disponível em: <https://www.britishcouncil.org.br/sites/default/files/demandas_de_aprendizagempesquisacompleta.pdf>. Acesso em: 25/02/2019.

BRITO, D. M. D. *et al*. Predição de desempenho de alunos do primeiro período baseado nas notas de ingresso utilizando métodos de aprendizagem de máquina. **XXV Simpósio Brasileiro de Informática na Educação (SBIE 2014)** , 2014. 882 - 890.

BRITTO NETO, Laurindo de Sousa. **Sistemas Wearable baseados em Métodos de Visão Computacional para auxiliar Pessoas com Deficiência Visual**. Campinas, SP, 2016. Originalmente apresentada como Tese (Doutorado) ao Instituto de Computação da Universidade Estadual de Campinas, 2016.

BUDOLA, T. Comissão de Educação ouve demandas do Centro de Línguas Estrangeiras Modernas. **Assembleia Legislativa do Estado do Paraná**, 2017. Disponível em: <http://www.assembleia.pr.leg.br/divulgacao/noticias/comissao-de-educacao-ouve-demandas-do-centro-de-linguas-estrangeiras-modernas>. Acesso em: 28 abril 2019.

CAMARGO, L. H. D. Turnover: machine learning reduz taxa em até 8% no. **SA Varejo**, 2018. Disponivel em: < https://www.savarejo.com.br/detalhe/ simples-assim-sa/turnover-machine-learning-reduz-taxa-em-ate-8-no-varejo >. Acesso em: 26 mar. 2019.

CARVALHO, Hialo Muniz. **Aprendizado de Máquina voltado para Mineração de Dados:** Árvores de Decisão, Brasília, DF, 2014. Originalmente apresentada como monografia ao curso de graduação em (Engenharia de Software) da Universidade de Brasília, 2014.

CHAN, Kok Wai; TEYMOURZADEH, Rozita; WAIDHUBA, Martin; VEE HOONG, Mok. (2013). **Smart Analytical Signature Verification For DSP Applications**. Proceedings - 2013 IEEE Conference on Systems, Process and Control, ICSPC 2013. 10.1109/SPC.2013.6735151.

CORRÊA, E. M. D.; MACHADO, J. A.; MELO, P. G. S. D. Estratégias de atração e relacionamento com os clientes na agência Boa Vista dos correios-RR. **Simpósio de Excelência em Gestão e Tecnologia (SEGET)**, Resende, 31 Outubro 2016.

CORTES, C.; VAPNIK, V. Support-vector networks. **Machine Learning**, v. 20, n. 3, p. 273-297, Setembro 1995. ISSN 1573-0565.

COVER, T. M.; HART, P. E. **Nearest Neighbor Pattern Classification, 2018***.* Disponível em: < https://www.cs.bgu.ac.il/~adsmb182/wiki.files/borak-lecture%20notes.pdf>. Acessado em: 03 abril 2019.

DATA SCIENCE ACADEMY. **17 Casos de Uso de Machine Learning***, 2018.* Disponível em: < http://datascienceacademy.com.br/blog/17-casos-de-uso-de-machine-learning/>. Acessado em: 26/ mar. 2018.

DIAS, M. F. R.; PASCUTTI, P. G.; SILVA, M. L. D. Aprendizado de Máquina e suas Aplicações em Bioinformática Machine Learning and Applications in Bioinformatics.

Revista Semioses, v. 1, n. 1, 2016.< http://dx.doi.org/10.15202/10.15202/1981-996X.2016v10n1p23>. Acessado em: 30 mar. 2019.

DOMINGOS, P. A Few Useful Things to Know About Machine Learning. **Communications of the ACM**, New York, v. 55, n. 10, p. 78-87, Outubro 2015.

DONGES, N. The Random Forest Algorithm. **Machine Learning - Blog**, 2018. Disponivel em: <https://machinelearning-blog.com/2018/02/06/the-random-forest-algorithm/>. Acesso em: 08 abril 2019.

EBRAHIMI, E.; MOLLAZADE, K.; ARMAN AREFI. An Expert System for Classification of Potato. **International Journal of Food Engineering**, v. 8, n. 9, jan. 2012.

FACELI, K. *et al*. **Inteligência Artificial:** Uma Abordagem de Aprendizado de Máquina. Rio de Janeiro: LTC, 2011

FISCHER, Frida Marina *et al*. Efeitos do trabalho sobre a saúde de adolescentes. **Ciência & saúde coletiva**. V.8, n.4, pp.973-984, 2003. ISSN 1413-8123.

FONSECA, T. M. M. *Ensinar* – Aprender, pensando a prática pedagógica. **Secretaria de Estado da Educação Superintendência da Educação Programa de Desenvolvimento Educacional – PDE** 2008. Disponível em: < http://www.diaadiaeducacao.pr.gov.br/portals/pde/arquivos/1782-6.pdf>. Acessado em: 21 nov 2018.

FREIRE, Rose Hélida Astolfo. **Possíveis Causas da Evasão Escolar e de Retorno na Educação de Jovens e Adultos.** Medianeira, 2014. Originalmente apresentada como monografia de Pesquisa e Pós-Graduação Especialização em Educação: Métodos e Técnicas de Ensino – Universidade Tecnológica Federal do Paraná, 2014.

GESTÃO DE ESTUDOS E AVALIAÇÃO DE INICIATIVAS PÚBLICAS (GESTA), 2017. **Políticas públicas para redução do abandono e evasão escolar de jovens**. Disponível em: <http://gesta.org.br/wp-content/uploads/2017/09/Politicas-Publicas-para-reducao-do-abandono-e-evasao-escolar-de-jovens.pdf>. Acessado em: 04 nov 2018

SILVEIRA, D. Metade dos trabalhadores brasileiros tem renda menor que o salário mínimo, aponta IBGE. **Globo**, 2017. Disponivel em: <https://g1.globo.com/economia/noticia/metade-dos-trabalhadores-brasileiros-tem-renda-menor-que-o-salario-minimo-aponta-ibge.ghtml>. Acesso em: 30 Junho 2019.

GOMES, Alberto Albuquerque. **Evasão e evadidos: o discurso dos ex-alunos sobre evasão escolar nos cursos de licenciatura**. 1998. 203 f. Tese (doutorado) - Universidade Estadual Paulista, Faculdade de Filosofia e Ciências, 1998. Disponível em: <http://hdl.handle.net/11449/102247>.

GONÇALVES G. S.; BARREIROS M. O.; BARREIROS S. P. O.; CORREIA L. C. Análise dos Fatores que Causam Dificuldades de Aprendizagem da Leitura e Escrita

nas Séries Iniciais do Ensino Fundamental. **Revista Espacios** Vol. 38 (Nº 60) Ano 2017. Pág. 11.

HONGYU, K.; SANDANIELO, V. L. M.; JUNIOR, G. J. D. O. Análise de Componentes Principais: resumo teórico, aplicação e interpretação. **Engineering and Science**, v. 1, n. 5, p. 83 - 90, 2015. ISSN 2358-5390

HOSMER, D. W, LEMESHOW S. Applied Logistic Regression. New York: **Wiley-Interscience**; 2000. 2ª Ed.

HOTZA, M. A. S. **O abandono nos cursos de graduação da UFSC em 1997: a percepção dos alunos-abandono**. Florianópolis,2000. Originalmente apresentado como Dissertação (Mestrado em Psicologia) - Universidade Federal de Santa Catarina, 2000.

JABBAR, H. K.; KHAN, R. Z. Methods to avoid over-fitting and under-fitting in supervised machine learning (comparative study), **Computer Science, Communication and Instrumentation Devices**, 2015.

JAMES G., WITTEN D.; HASTIE T., TIBSHIRANI R. **An Introduction to Statistical Learning with Applications in R**. v.6, p. 33, 2015.

KANTORSKI, G. Z. et al. Uma visão do futuro: Previsão de evasão em cursos de graduação presenciais de universidades públicas: o caso do curso de zootecnia. **XV Colóquio internacional de gestão universitária – CIGU,** Mal del Plata, 2 dezembro 2015.

KOTLER, P.; KELLER. K. L. **Administração de Marketing**. São Paulo: Pearson Education do Brasil, n. 14, 2012. ISBN 978-85-8143-000-3

KOTSIANTIS, S. B. **Supervised Machine Learning: A Review of Classification Technique**, p. 249–268 Disponível em: <http://www.informatica.si/index.php/informatica/article/viewFile/148/140>. Acessado em: 02 abril 2018.

LOPES, R. S. L. **A relação professor aluno e o processo ensino aprendizagem**, 2011 Disponível em: <http://www.diaadiaeducacao.pr.gov.br/portals/pde/arquivos/1534-8.pdf.>. Acessado em: 03 nov. 2018.

MARTINS, Ana Natacha. **Trabalho de Investigação Sobre a Evasão nos Cursos de Inglês da UFRGS.** Porto Alegre, 2011. Originalmente apresentada como monografia de Graduação do Instituto de Letras da Universidade Federal do Rio Grande do Sul, 2011.

MITCHELL, T. **Machine Learning. New York**, NY: McGraw-Hill, 1997. ISBN0-07-042807-7.

MOHAMMED, M.; PATHAN, A.-S. K. **Automatic Defense Against Zero-day Polymorphic Worms in Communication Networks**. Boca Raton: CRC Press, 2013. ISBN 9781482219050.

MONTEIRO, Rodrigo Bezerra. **Comparação de Técnicas de Aprendizado de Máquina para Predição da Disponibilidade de Bicicletas no Projeto Bicicletar Fortaleza**. Quixadá, 2018. Originalmente apresentado como o em Sistemas de Informação do Campus Quixadá da Universidade Federal do Ceará, 2018.

NERI, M. **Tempo de Permanência na Escola**, Rio de Janeiro. FGV/IBRE, CPS, 2009. Disponível em: <https://www.cps.fgv.br/ibrecps/rede/tpe/>. Acesso em: 30 Maio 2019.

NERI, M., 2010. **Motivos da evasão escolar.** Disponível em: <http://www.cps.fgv.br/ibrecps/TPE/TPE_MotivacoesEscolares_fim.pdf>. Acessado em: 21 nov. 2018

NEVES, Samuel Antônio. **Técnicas de Aprendizado de Máquina Aplicadas a Classificação da Qualidade de Pavimentos Asfálticos utilizando Smartphones.** João Monlevade, 2018. Originalmente apresentado como monografia o de Engenharia de Computação do Instituto de Ciências Exatas e Aplicadas, da Universidade Federal de Ouro Preto, 2018.

OLIVEIRA JÚNIOR, José Gonçalves. **Identificação de Padrões para a Análise da Evasão em Cursos de Graduação Usando Mineração de Dados Educacionais**. 2015. 86 f. Dissertação - Programa de Pós-graduação em Computação Aplicada, Universidade Tecnológica Federal do Paraná. Curitiba, 2015.

PEREIRA D. R.. BOTELHO M. A. DA S. **Satisfação e Fidelização no Ensino Superior: Um Estudo de Correlação em uma IES Privada de Belém-PA**, 2009. Disponível em: <https://www.aedb.br/seget/arquivos/artigos09/192_ artigoidentsegetgu.pdf>. Acessado em: 22 nov. 2018.

PINHEIRO, M. A. L.; SILVA, J. C. D.; , B. F. D. S. Aprendizado de Maquina Aplicado à Análise de Evasão no Ensino Superior. **IX Computer on the Beach**, Florianópolis, 2018. 512-521.

RADIYA-DIXIT, E.; ZHU, D.; BECK, A. H. Automated Classification of Benign and Malignant Proliferative Breast Lesions. **Scientific Reports**, v. 7, n. 1, dez. 2017.

ROBBINS. S. P. **Comportamento organizacional, tradução técnica Reynaldo Marcondes**, 11. ed.- São Paulo: Pearson Prentice Hall, 2005.

ROZA, Felippe Schmoeller. **Aprendizagem de máquina para apoio à tomada de decisão em vendas do varejo utilizando registros de vendas**, Florianópolis, 2016. Originalmente apresentado como monografia de graduação do Curso de Engenharia de Controle e Automação para Universidade Federal de Santa Catarina.

SHALEV-SHWARTZ, S.; BEN-DAVID, S. **Understanding Machine Learning**: From Theory to Algorithms. New York: Cambridge University Press., 2014. ISBN 978-1-107-05713-5.

SIDOSKI, Eliane. Estudo da demanda por profissionais bilíngues no mercado turístico do município de Fernandes Pinheiro – PR., Irati, 2014. Originalmente apresentado como monografia para conclusão do curso da Universidade Estadual do Centro-Oeste, campus de Irati.

SILVEIRA, A. L. F. *et al*. Congresso Internacional,ABED de Educação A Distância (21º CIAED). **Análise de Perfil e Necessidades dos Alunos:** Fatores Determinantes para a Mediação da Aprendizagem na Tutoria da EAD Sebrae, Florianópolis , 2015.

SMOLA, A.; VISHWANATHAN, S. **Introduction to machine learning**. Cambridge University, UK, v. 32, p. 10, 2008.

SOARES, T. M. *et al*. Fatores associados ao abandono escolar no ensino médio público de Minas Gerais. **Educação e Pesquisa**, São Paulo, v. 41, n. 3, p. 757-772, Julho 2015.

SOUSA, Erica Soares Brito. **Evasão em um curso de inglês:** um estudo exploratório de suas principais causas. Pedro Leopoldo, 2008. Originalmente apresentada com dissertação (Mestre) ao Curso de Mestrado Profissional em Administração das Faculdades Integradas de Pedro Leopoldo

STEARNS, E.; GLENNIE, E. J. When and Why Dropouts Leave High School. **Youth & Society**, v. 38, n. 1, p. 29-57, setembro 2006. Doi: 10.1177/0044118X05282764.

TARCA, A. L. *et al*. Machine Learning and Its Applications. **PLoS Computational Biology**, v. 3, n. 6, p. 953- 963, junho 2007.

TONDELLI, Maria de Fátima**. A Influência da língua estrangeira na empregabilidade de profissionais da área tecnológica no setor industrial:** um estudo exploratório na região norte do Paraná, Ponta Grossa, 2005. Originalmente apresentado como dissertação de Mestre em Engenharia de Produção, do Programa de Pós-Graduação em Engenharia de Produção para a Universidade Tecnológica Federal do Paraná, campus Ponta Grossa.

VERGARA, S. C. **Gestão de pessoas**. 8. ed. São Paulo: Atlas, 2009.

VERIKAS, A. *et al*. Electromyographic Patterns during Golf Swing: Activation Sequence Profiling and Prediction of Shot Effectiveness. **Sensors**, v. 16, n. 4, 2016. ISSN 1424-8220.

WITTEN, I. H..; FRANK, E. **Data Mining:** Practical Machine Learning Tools and Techniques. 2ª. ed. San Francisco: Elsevier, 2005.

ZIMMER, Gessi Terezinha. **Percepções sobre a Evasão no Primeiro Ano do Ensino Médio em uma Escola Pública de Sarandi (RS)**, Sarandi, 2013.

Originalmente apresentado como monografia ao Curso de Pós-Graduação a Distância Especialização Lato Sensu em Gestão Educacional, da Universidade Federal de Santa Maria (UFSM, RS), 2013.

APÊNDICE A - Questionário Caso 1 e Caso 2

PESQUISA SOBRE O PERFIL DO ESTUDANTE DE LÍNGUAS ESTRANGEIRAS

O objetivo desta pesquisa é identificar o perfil dos indivíduos que estudam ou já estudaram línguas estrangeiras.

Tempo esperado de preenchimento: 5 minutos

TERMO DE CONSENTIMENTO LIVRE E ESCLARECIDO (TCLE)

Este questionário se trata de uma pesquisa conduzida por Monique Tamara de Lima (R.A 1838831) discente do Curso de Engenharia de Produção da Universidade Tecnológica Federal do Paraná (UTFPR) - Câmpus Londrina, sob a supervisão do Professor Doutor Rafael Henrique Palma Lima. Tendo como enfoque prever a evasão acadêmica de cursos de idiomas cujo objetivo é melhorar a experiência de aprendizado dos estudantes e reduzir a taxa de evasão, por meio da identificação de atributos e padrões que ocasionam esse evento.

Para isto estamos a pedir-lhe a sua colaboração na participação voluntária nesta pesquisa, respondendo este questionário. A sua participação é de grande importância para o sucesso desta investigação. Todavia, há a opção de não participar, ou até mesmo interromper a qualquer momento o preenchimento do questionário, sem nenhum prejuízo ou coação. Todas as informações obtidas nesse estudo são de natureza confidencial e de total anonimato dos participantes. Os resultados serão apresentados de maneira agregada publicações cientificas de cunho acadêmico.

( ) Concordo ( ) Discordo

2. Em que situação você se enquadra neste questionário?

( ) Caso 1 - Está cursando atualmente **1** curso de idioma
( ) Caso 2 - Está cursando atualmente **mais de 1** curso de idioma
( ) Caso 3 - Já fez curso de idioma, mas atualmente não está cursando
( ) Caso 4 - Já concluiu o curso de idioma, e atualmente não está cursando nenhum outro curso de línguas estrangeiras
( ) Caso 5 – Nenhum dos casos anteriores

Caso tenha realizado mais de um caso responda o questionário apenas sobre 1 curso de idioma, escolhendo o curso que você julga ser o mais importante!!!

Perfil do respondente

3. Qual a sua idade em anos? _____

4. Renda familiar

( ) Até 2 salários mínimos
( ) Entre 2 e 4 salários mínimos
( ) Entre 4 e 6 salários mínimos
( ) Entre 6 e 10 salários mínimos
( ) Entre 10 a 20 salários mínimos
( ) acima de 20 salários mínimos

5. Sexo

( ) Feminino ( ) Masculino ( ) Prefiro não responder

6. Estado Civil

( ) Solteiro(a) ( ) Casado(a) ( ) Separado(a)/desquitado(a)/divorciado(a) ( ) Viúvo(a) ( ) Outro

7. Quantos filhos você possui?

( ) 0 – Nenhum ( ) 1 filho ( ) 2 filhos ( ) 3 filhos ( ) 4 ou mais filhos

8. Qual o seu nível de escolaridade?
( ) Ensino Fundamental – Incompleto ( ) Ensino Fundamental – Completo ( ) Ensino Médio – Incompleto ( ) Ensino Médio Completo
( ) Ensino Superior - Incompleto ( ) Ensino Superior Completo ( ) Outro:

9. Qual o nível de escolaridade de sua **mãe**?
( ) Sem escolaridade ( ) Ensino Fundamental – Incompleto ( ) Ensino Fundamental - Completo ( ) Ensino Médio – Incompleto
( ) Ensino Médio – Completo ( ) Ensino Superior – Incompleto ( ) Mestrado ou Doutorado ( ) Não sei

10. Qual o nível de escolaridade de seu **pai**?
( ) Sem escolaridade ( ) Ensino Fundamental – Incompleto ( ) Ensino Fundamental - Completo ( ) Ensino Médio – Incompleto
( ) Ensino Médio – Completo ( ) Ensino Superior – Incompleto ( ) Mestrado ou Doutorado ( ) Não sei

Questões relativas ao Contexto do Curso

11. Você está respondendo o questionário sobre qual idioma?
( ) Inglês ( ) Espanhol ( ) Francês ( ) Italiano ( ) Italiano
( ) Alemão ( ) Japonês ( ) Outros:

12. Sobre a frase: "Há possibilidade de eu desistir do curso no próximo semestre"
( ) Discordo Totalmente ( ) Discordo ( ) Nem concordo, nem discordo ( ) Concordo ( ) Concordo Totalmente

13. Há quanto tempo você está estudando este curso de idioma?
( ) até 6 meses ( ) de 6 meses a 1 ano ( ) de 1 a 2 anos ( ) de 2 a 3 anos ( ) mais de 3 anos

14. Qual é a duração total do Curso?
( ) até 1 ano ( ) de 1 a 2 anos ( ) de 2 a 3 anos ( ) 3 a 4 anos ( ) mais de 4 anos

15. Quanto tempo falta para que você se forme em seu curso atual?
( ) até 6 meses ( ) de 6 meses a 1 ano ( ) de 1 a 2 anos ( ) de 2 a 3 anos ( ) mais de 3 anos

16. Qual seu grau de concordância com a seguinte frase: "Eu tenho intenção de concluir este curso de idioma, cursando todas as etapas previstas".
( ) Discordo Totalmente ( ) Discordo ( ) Nem concordo, nem discordo ( ) Concordo ( ) Concordo Totalmente

17. Como você considera o valor da mensalidade do curso em relação a sua renda mensal?
( ) Muito caro ( ) Caro ( ) Nem caro, nem barato ( ) Barato ( ) Insignificante

18. Assinale alternativas que relacionem situações que ocorreram durante a realização do curso (é permitido assinalar múltiplas alternativas)

( ) Separou-se/divorciou-se
( ) Mudou-se de endereço
( ) Teve filho(s)
( ) Outro:_____
( ) Começou a trabalhar
( ) Mudou o idioma estudado
( ) Realizou viagens
( ) Mudou-se de instituição de ensino
( ) Perdeu o emprego
( ) Enfrentou problemas de saúde
( ) Nenhuma das alternativas

19. O curso é:
( ) Presencial ( ) À distância

Dificuldades dos Cursos PRESENCIAIS

(apenas para quem faz cursos presenciais)

20. Como você avalia a distância a ser percorrida até a instituição de ensino?
( ) Muito perto ( ) Perto ( ) Nem longe, nem perto ( ) Longe ( ) Muito longe

21. Como você avalia a facilidade de transporte até a instituição de ensino?
( ) Muito fácil ( ) Fácil ( ) Nem difícil, nem fácil ( ) Difícil ( ) Muito Difícil

22. Como você vai para o curso?
( ) Carona ( ) Bicicleta ( ) Motocicleta própria ( ) Carro próprio ( ) Transporte público
( ) Caminhando ( ) Outro:_____

Dificuldades dos Cursos à DISTÂNCIAS

(apenas para quem faz cursos à distâncias)

20. Assinale o grau de concordância com a seguinte frase: “Eu tenho facilidade para dominar novas tecnologias, sendo assim não tenho dificuldades para interagir com o ambiente virtual de aprendizagem”.

( ) Discordo Totalmente ( ) Discordo ( ) Nem concordo, nem discordo ( ) Concordo ( ) Concordo Totalmente

21. Assinale o grau de concordância com a seguinte frase: “Eu não enfrento dificuldades com o material didático nem tenho problemas de comunicação com o tutor, ele sempre me estimula e fornece feedback sobre o meu desempenho”.

( ) Discordo Totalmente ( ) Discordo ( ) Nem concordo, nem discordo ( ) Concordo ( ) Concordo Totalmente

Questões relativas ao Nível de Satisfação - Curso Presencial

(apenas para aqueles que fizeram cursos presenciais)

23. Como você classifica a sua **motivação** para o estudo de idiomas em geral:

( ) Totalmente desmotivado ( ) Desmotivado ( ) Nem motivado, nem desmotivado ( ) Motivado ( )Muito Motivado

24. Nível de satisfação com o **idioma escolhido**:

( ) Muito insatisfeito ( ) Insatisfeito ( ) Nem satisfeito, nem insatisfeito ( ) Satisfeito ( ) Muito satisfeito

25. De acordo com sua percepção, o quanto é relevante o aprendizado do idioma escolhido para a sua **carreira**?

( ) Sem importância ( ) Pouco importante ( ) Importante ( ) Muito importante ( ) Fundamental

26. Qual a importância de aprender o idioma escolhido para a sua **vida pessoal**?

( ) Sem importância ( ) Pouco importante ( ) Importante ( ) Muito importante ( ) Fundamental

27. Como você avalia seu nível de dificuldade para aprender idiomas?

( ) Muito baixa ( ) baixa ( ) Nem baixa, nem alta ( ) alta ( ) Muito alta

28. Quando comparado o seu nível de domínio do idioma estudado com o nível dos demais alunos, você acredita que os estudantes estão:

( ) Muito abaixo do seu nível ( ) Abaixo do seu nível ( ) Mesmo nível ( ) Acima do seu nível ( ) Muito acima do seu nível

29. Qual é o seu grau de dificuldade para conciliar a carga horária do curso com suas demais atividades do dia a dia?

( ) Muita baixo ( ) Baixo ( ) Nem alto, nem baixo ( ) Alto ( ) Muito alto

30. Como você avalia os horários das aulas?

( ) Péssimo ( ) Ruim ( ) Nem bom, nem ruim ( ) Bom ( ) Ótimo

31. Como é a sua integração com os demais alunos da classe?

( ) Péssimo ( ) Ruim ( ) Nem bom, nem ruim ( ) Bom ( ) Ótimo

32. Como você classifica o seu relacionamento com o professor?

( ) Péssimo ( ) Ruim ( ) Nem bom, nem ruim ( ) Bom ( ) Ótimo

33. Qual o seu nível de satisfação com a instituição de ensino?

( ) Muito insatisfeito ( ) Insatisfeito ( ) Nem satisfeito, nem insatisfeito ( ) Satisfeito ( ) Muito satisfeito

34. Você poderia nos fornece o seu e-mail para receber os resultados obtidos e para realizarmos uma segunda etapa desta pesquisa? (Opcional)

_____

35. Caso você tenha alguma reclamação, sugestão de melhoria, dúvidas, ou feedback, por gentileza, relate-o no campo abaixo. (Opcional)

_____

Questões relativas ao Nível de Satisfação – Curso à Distância

(apenas para aqueles que fizeram cursos à distância)

22. Como você classifica a sua **motivação** para o estudo de idiomas em geral:

( ) Totalmente desmotivado ( ) Desmotivado ( ) Nem motivado, nem desmotivado ( ) Motivado ( ) Muito Motivado

23. Nível de satisfação com o **idioma escolhido**:

( ) Muito insatisfeito ( ) Insatisfeito ( ) Nem satisfeito, nem insatisfeito ( ) Satisfeito ( ) Muito satisfeito

24. De acordo com sua percepção, o quanto é relevante o aprendizado do idioma escolhido para a sua **carreira**?

( ) Sem importância ( ) Pouco importante ( ) Importante ( ) Muito importante ( ) Fundamental

25. Qual a importância de aprender o idioma escolhido para a sua **vida pessoal**?

( ) Sem importância ( ) Pouco importante ( ) Importante ( ) Muito importante ( ) Fundamental

26. Como você avalia seu nível de dificuldade para aprender idiomas?

( ) Muito baixa ( ) Baixa ( ) Nem baixa, nem alta ( ) Alta ( ) Muito Alta

27. Quando comparado o nível de exigência do módulo que você esteja cursando, com o seu nível de domínio do idioma estudado, você acredita que o módulo está:

( ) Muito abaixo do seu nível ( ) Abaixo do seu nível ( ) Mesmo nível ( ) Acima do seu nível ( ) Muito acima do seu nível

28. Qual é o seu grau de dificuldade para conciliar a carga horária do curso com suas demais atividades do dia a dia?

( ) Muito baixo ( ) Baixo ( ) Nem alto, nem baixo ( ) Alto ( ) Muito alto

29. Como você avalia a plataforma e suas funcionalidades do curso?

( ) Péssimo ( ) Ruim ( ) Nem bom, nem ruim ( ) Bom ( ) Ótimo

30. Como você avalia a metodologia do idioma estudado?

( ) Péssimo ( ) Ruim ( ) Nem bom, nem ruim ( ) Bom ( ) Ótimo

31. Como você avalia a sua disciplina para realizar as atividades propostas no curso de idioma.

( ) Péssima ( ) Ruim ( ) Nem bom, nem ruim ( ) Bom ( ) Ótima

32. Qual o seu nível de satisfação com a instituição de ensino?

( ) Muito insatisfeito ( ) Insatisfeito ( ) Nem satisfeito, nem insatisfeito ( ) Satisfeito ( ) Muito satisfeito

33. Você poderia nos fornece o seu e-mail para receber os resultados obtidos e para realizarmos uma segunda etapa desta pesquisa? *(Opcional)*

_____

34. Caso você tenha alguma reclamação, sugestão de melhoria, dúvidas, ou feedback, por gentileza, relate-o no campo abaixo. *(Opcional)*

_____

APÊNDICE B - Questionário Caso 3

PESQUISA SOBRE O PERFIL DO ESTUDANTE DE LÍNGUAS ESTRANGEIRAS

O objetivo desta pesquisa é identificar o perfil dos indivíduos que estudam ou já estudaram línguas estrangeiras.

Tempo esperado de preenchimento: 5 minutos

TERMO DE CONSENTIMENTO LIVRE E ESCLARECIDO (TCLE)

Este questionário se trata de uma pesquisa conduzida por Monique Tamara de Lima (R.A 1838831) discente do Curso de Engenharia de Produção da Universidade Tecnológica Federal do Paraná (UTFPR) - Câmpus Londrina, sob a supervisão do Professor Doutor Rafael Henrique Palma Lima. Tendo como enfoque prever a evasão acadêmica de cursos de idiomas cujo objetivo é melhorar a experiência de aprendizado dos estudantes e reduzir a taxa de evasão, por meio da identificação de atributos e padrões que ocasionam esse evento.

Para isto estamos a pedir-lhe a sua colaboração na participação voluntária nesta pesquisa, respondendo este questionário. A sua participação é de grande importância para o sucesso desta investigação. Todavia, há a opção de não participar, ou até mesmo interromper a qualquer momento o preenchimento do questionário, sem nenhum prejuízo ou coação. Todas as informações obtidas nesse estudo são de natureza confidencial e de total anonimato dos participantes. Os resultados serão apresentados de maneira agregada publicações cientificas de cunho acadêmico.

( ) Concordo ( ) Discordo

2. Em que situação você se enquadra neste questionário?

( ) Caso 1 - Está cursando atualmente **1** curso de idioma
( ) Caso 2 - Está cursando atualmente **mais de 1** curso de idioma
( ) Caso 3 - Já fez curso de idioma, mas atualmente não está cursando
( ) Caso 4 - Já concluiu o curso de idioma, e atualmente não está cursando nenhum outro curso de línguas estrangeiras
( ) Caso 5 – Nenhum dos casos anteriores

Caso tenha realizado mais de um caso responda o questionário apenas sobre 1 curso de idioma, escolhendo o curso que você julga ser o mais importante!!!

Perfil do respondente

3. Qual a sua idade em anos? _____

4. Renda familiar

( ) Até 2 salários mínimos
( ) Entre 2 e 4 salários mínimos
( ) Entre 4 e 6 salários mínimos
( ) Entre 6 e 10 salários mínimos
( ) Entre 10 a 20 salários mínimos
( ) acima de 20 salários mínimos

5. Sexo

( ) Feminino ( ) Masculino ( ) Prefiro não responder

6. Estado Civil

( ) Solteiro(a) ( ) Casado(a) ( ) Separado(a)/desquitado(a)/divorciado(a) ( ) Viúvo(a) ( ) Outro

7. Quantos filhos você possui?
( ) 0 – Nenhum ( ) 1 filho ( ) 2 filhos ( ) 3 filhos ( ) 4 ou mais filhos

8. Qual o seu nível de escolaridade?
( ) Ensino Fundamental – Incompleto ( ) Ensino Fundamental – Completo ( ) Ensino Médio – Incompleto ( ) Ensino Médio Completo
( ) Ensino Superior - Incompleto ( ) Ensino Superior Completo ( ) Outro:

9. Qual o nível de escolaridade de sua **mãe**?
( ) Sem escolaridade ( ) Ensino Fundamental – Incompleto ( ) Ensino Fundamental - Completo ( ) Ensino Médio – Incompleto
( ) Ensino Médio – Completo ( ) Ensino Superior – Incompleto ( ) Mestrado ou Doutorado ( ) Não sei

10. Qual o nível de escolaridade de seu **pai**?
( ) Sem escolaridade ( ) Ensino Fundamental – Incompleto ( ) Ensino Fundamental - Completo ( ) Ensino Médio – Incompleto
( ) Ensino Médio – Completo ( ) Ensino Superior – Incompleto ( ) Mestrado ou Doutorado ( ) Não sei

Questões relativas ao Contexto do Curso

11. Você está respondendo o questionário sobre qual idioma?
( ) Inglês ( ) Espanhol ( ) Francês ( ) Italiano ( ) Italiano
( ) Alemão ( ) Japonês ( ) Outros:

12. Quantos anos você tinha quando desistiu do curso? _____

13. Por quanto tempo você estudou este curso de idioma?
( ) até 6 meses ( ) de 6 meses a 1 ano ( ) de 1 a 2 anos ( ) de 2 a 3 anos ( ) mais de 3 anos

14. Qual é a duração total do Curso?
( ) até 1 ano ( ) de 1 a 2 anos ( ) de 2 a 3 anos ( ) 3 a 4 anos ( ) mais de 4 anos

15. Quanto tempo faltava para que você se formasse em seu curso de idioma?
( ) até 6 meses ( ) de 6 meses a 1 ano ( ) de 1 a 2 anos ( ) de 2 a 3 anos ( ) mais de 3 anos

16. Qual seu grau de concordância com a seguinte frase: “Eu tenho intenção de voltar a fazer o curso para concluí-lo, cursando todas as etapas previstas”
( ) Discordo Totalmente ( ) Discordo ( ) Nem concordo, nem discordo ( ) Concordo ( ) Concordo Totalmente

17. Sobre a frase: “Há possibilidade de voltar a fazer o curso de idioma no próximo semestre”.
( ) Discordo Totalmente ( ) Discordo ( ) Nem concordo, nem discordo ( ) Concordo ( ) Concordo Totalmente

18. Como você avaliava o valor da mensalidade do curso em relação a sua renda mensal?
( ) Muito caro ( ) Caro ( ) Nem caro, nem barato ( ) Barato ( ) Insignificante

19. Assinale alternativas que relacionem situações que ocorreram durante a realização do curso (é permitido assinalar múltiplas alternativas)

( ) Separou-se/divorciou-se
( ) Mudou-se de endereço
( ) Teve filho(s)
( ) Outro:_____
( ) Começou a trabalhar
( ) Mudou o idioma estudado
( ) Realizou viagens
( ) Mudou-se de instituição de ensino
( ) Perdeu o emprego
( ) Enfrentou problemas de saúde
( ) Nenhuma das alternativas

20. O curso era:
( ) Presencial ( ) À distância

Dificuldades dos Cursos PRESENCIAIS

(apenas para quem faz cursos presenciais)

21. Como você avaliava a distância a ser percorrida até a instituição de ensino?
( ) Muito perto ( ) Perto ( ) Nem longe, nem perto ( ) Longe ( ) Muito longe

22. Como você avaliava a facilidade de transporte até a instituição de ensino?
( ) Muito fácil ( ) Fácil ( ) Nem difícil, nem fácil ( ) Difícil ( ) Muito Difícil

23. Como você ia para o curso?
( ) Carona ( ) Bicicleta ( ) Motocicleta própria ( ) Carro próprio ( ) Transporte público

Dificuldades dos Cursos à DISTÂNCIAS
(apenas para quem faz cursos à distâncias)

21. Assinale o grau de concordância com a seguinte frase: "Eu tinha facilidade para dominar novas tecnologias, sendo assim não tive dificuldades para interagir com o ambiente virtual de aprendizagem".
( ) Discordo Totalmente ( ) Discordo ( ) Nem concordo, nem discordo ( ) Concordo ( ) Concordo Totalmente

22. Assinale o grau de concordância com a seguinte frase: "Eu não enfrentei dificuldades com o material didático nem tive problemas de comunicação com o tutor, ele sempre me estimulava e fornecia feedback sobre o meu desempenho".
( ) Discordo Totalmente ( ) Discordo ( ) Nem concordo, nem discordo ( ) Concordo ( ) Concordo Totalmente

Questões relativas ao Nível de Satisfação - Curso Presencial
(apenas para aqueles que fizeram cursos presenciais)
RESPONDA AS QUESTÕES ABAIXO, CONSIDERANDO A PERCEPÇÃO QUE VOCÊ TINHA DURANTE A REALIZAÇÃO DO CURSO, NÃO A SUA PERCEPÇÃO ATUAL.

24. Como você classificava a sua **motivação** para o estudo de idiomas em geral:
( ) Totalmente desmotivado ( ) Desmotivado ( ) Nem motivado, nem desmotivado ( ) Motivado ( )Muito Motivado

25. Qual era o seu nível de satisfação com o **idioma escolhido**:
( ) Muito insatisfeito ( ) Insatisfeito ( ) Nem satisfeito, nem insatisfeito ( ) Satisfeito ( ) Muito satisfeito

26. De acordo com sua percepção, o quanto você considerava relevante o aprendizado do idioma escolhido para a sua **carreira**?
( ) Sem importância ( ) Pouco importante ( ) Importante ( ) Muito importante ( ) Fundamental

27. Como você classificava a importância de aprender o idioma escolhido para a sua **vida pessoal**?
( ) Sem importância ( ) Pouco importante ( ) Importante ( ) Muito importante ( ) Fundamental

28. Como você avaliava seu nível de dificuldade em aprender idiomas?
( ) Muito baixa ( ) baixa ( ) Nem baixa, nem alta ( ) alta ( ) Muito alta

29. Quando era comparado o seu nível de domínio do idioma estudado com o nível dos demais alunos, você acreditava que os estudantes estavam:
( ) Muito abaixo do seu nível ( ) Abaixo do seu nível ( ) Mesmo nível ( ) Acima do seu nível ( ) Muito acima do seu nível

30. Qual era o seu grau de dificuldade para conciliar a carga horária do curso com suas demais atividades do dia a dia?
( ) Muita baixo ( ) Baixo ( ) Nem alto, nem baixo ( ) Alto ( ) Muito alto

31. Como você avaliava os horários das aulas?
( ) Péssimo ( ) Ruim ( ) Nem bom, nem ruim ( ) Bom ( ) Ótimo

32. Como era a sua integração com os demais alunos da classe?
( ) Péssimo ( ) Ruim ( ) Nem bom, nem ruim ( ) Bom ( ) Ótimo

33. Como você classificava o seu relacionamento com o professor?
( ) Péssimo ( ) Ruim ( ) Nem bom, nem ruim ( ) Bom ( ) Ótimo

34. Qual era o seu nível de satisfação com a instituição de ensino?
( ) Muito insatisfeito ( ) Insatisfeito ( ) Nem satisfeito, nem insatisfeito ( ) Satisfeito ( ) Muito satisfeito

35. Você poderia nos fornece o seu e-mail para receber os resultados obtidos e para realizarmos uma segunda etapa desta pesquisa? (Opcional)

_____

36. Caso você tenha alguma reclamação, sugestão de melhoria, dúvidas, ou feedback, por gentileza, relate-o no campo abaixo. (Opcional)

_____

Questões relativas ao Nível de Satisfação – Curso à Distância

RESPONDA AS QUESTÕES ABAIXO, CONSIDERANDO A PERCEPÇÃO QUE VOCÊ TINHA DURANTE A REALIZAÇÃO DO CURSO, NÃO A SUA PERCEPÇÃO ATUAL.

23. Como você classificava a sua **motivação** para o estudo de idiomas em geral::

( ) Totalmente desmotivado ( ) Desmotivado ( ) Nem motivado, nem desmotivado ( ) Motivado ( ) Muito Motivado

24. Qual era o seu nível de satisfação com o **idioma escolhido**:

( ) Muito insatisfeito ( ) Insatisfeito ( ) Nem satisfeito, nem insatisfeito ( ) Satisfeito ( ) Muito satisfeito

25. De acordo com sua percepção, o quanto você considerava relevante o aprendizado do idioma escolhido para a sua **carreira**?

( ) Sem importância ( ) Pouco importante ( ) Importante ( ) Muito importante ( ) Fundamental

26. Como você classificava a importância de aprender o idioma escolhido para a sua **vida pessoal**?

( ) Sem importância ( ) Pouco importante ( ) Importante ( ) Muito importante ( ) Fundamental

27. Como você avaliava seu nível de dificuldade em aprender idiomas?

( ) Muito baixa ( ) Baixa ( ) Nem baixa, nem alta ( ) Alta ( ) Muito Alta

28. Quando era comparado o nível de exigência do módulo que você estava cursando, com o seu nível de domínio do idioma estudado, você acredita que o módulo estava:

( ) Muito abaixo do seu nível ( ) Abaixo do seu nível ( ) Mesmo nível ( ) Acima do seu nível ( ) Muito acima do seu nível

29. Qual era o seu grau de dificuldade para conciliar a carga horária do curso com suas demais atividades do dia a dia?

( ) Muito baixo ( ) Baixo ( ) Nem alto, nem baixo ( ) Alto ( ) Muito alto

30. Como você avaliava a plataforma e suas funcionalidades do curso?

( ) Péssimo ( ) Ruim ( ) Nem bom, nem ruim ( ) Bom ( ) Ótimo

31. Como você avaliava a metodologia do idioma estudado?

( ) Péssimo ( ) Ruim ( ) Nem bom, nem ruim ( ) Bom ( ) Ótimo

32. Como você avalia a sua disciplina para realizar as atividades propostas no curso de idioma.

( ) Péssima ( ) Ruim ( ) Nem bom, nem ruim ( ) Bom ( ) Ótima

33. Qual o seu nível de satisfação com a instituição de ensino?

( ) Muito insatisfeito ( ) Insatisfeito ( ) Nem satisfeito, nem insatisfeito ( ) Satisfeito ( ) Muito satisfeito

34. Você poderia nos fornece o seu e-mail para receber os resultados obtidos e para realizarmos uma segunda etapa desta pesquisa? *(Opcional)*

_____

35. Caso você tenha alguma reclamação, sugestão de melhoria, dúvidas, ou feedback, por gentileza, relate-o no campo abaixo. *(Opcional)*

_____

APÊNDICE C - Questionário Caso 4

PESQUISA SOBRE O PERFIL DO ESTUDANTE DE LÍNGUAS ESTRANGEIRAS

O objetivo desta pesquisa é identificar o perfil dos indivíduos que estudam ou já estudaram línguas estrangeiras.

Tempo esperado de preenchimento: 5 minutos

TERMO DE CONSENTIMENTO LIVRE E ESCLARECIDO (TCLE)

Este questionário se trata de uma pesquisa conduzida por Monique Tamara de Lima (R.A 1838831) discente do Curso de Engenharia de Produção da Universidade Tecnológica Federal do Paraná (UTFPR) - Câmpus Londrina, sob a supervisão do Professor Doutor Rafael Henrique Palma Lima. Tendo como enfoque prever a evasão acadêmica de cursos de idiomas cujo objetivo é melhorar a experiência de aprendizado dos estudantes e reduzir a taxa de evasão, por meio da identificação de atributos e padrões que ocasionam esse evento.

Para isto estamos a pedir-lhe a sua colaboração na participação voluntária nesta pesquisa, respondendo este questionário. A sua participação é de grande importância para o sucesso desta investigação. Todavia, há a opção de não participar, ou até mesmo interromper a qualquer momento o preenchimento do questionário, sem nenhum prejuízo ou coação. Todas as informações obtidas nesse estudo são de natureza confidencial e de total anonimato dos participantes. Os resultados serão apresentados de maneira agregada publicações cientificas de cunho acadêmico.

( ) Concordo ( ) Discordo

2. Em que situação você se enquadra neste questionário?

( ) Caso 1 - Está cursando atualmente **1** curso de idioma
( ) Caso 2 - Está cursando atualmente **mais de 1** curso de idioma
( ) Caso 3 - Já fez curso de idioma, mas atualmente não está cursando
( ) Caso 4 - Já concluiu o curso de idioma, e atualmente não está cursando nenhum outro curso de línguas estrangeiras
( ) Caso 5 – Nenhum dos casos anteriores

Caso tenha realizado mais de um caso responda o questionário apenas sobre 1 curso de idioma, escolhendo o curso que você julga ser o mais importante!!!

Perfil do respondente

3. Qual a sua idade em anos? _____

4. Renda familiar

( ) Até 2 salários mínimos
( ) Entre 2 e 4 salários mínimos
( ) Entre 4 e 6 salários mínimos
( ) Entre 6 e 10 salários mínimos
( ) Entre 10 a 20 salários mínimos
( ) acima de 20 salários mínimos

5. Sexo

( ) Feminino ( ) Masculino ( ) Prefiro não responder

6. Estado Civil

( ) Solteiro(a) ( ) Casado(a) ( ) Separado(a)/desquitado(a)/divorciado(a) ( ) Viúvo(a) ( ) Outro

7. Quantos filhos você possui?
( ) 0 – Nenhum ( ) 1 filho ( ) 2 filhos ( ) 3 filhos ( ) 4 ou mais filhos

8. Qual o seu nível de escolaridade?
( ) Ensino Fundamental – Incompleto ( ) Ensino Fundamental – Completo ( ) Ensino Médio – Incompleto ( ) Ensino Médio Completo
( ) Ensino Superior - Incompleto ( ) Ensino Superior Completo ( ) Outro:

9. Qual o nível de escolaridade de sua **mãe**?
( ) Sem escolaridade ( ) Ensino Fundamental – Incompleto ( ) Ensino Fundamental - Completo ( ) Ensino Médio – Incompleto
( ) Ensino Médio – Completo ( ) Ensino Superior – Incompleto ( ) Mestrado ou Doutorado ( ) Não sei

10. Qual o nível de escolaridade de seu **pai**?
( ) Sem escolaridade ( ) Ensino Fundamental – Incompleto ( ) Ensino Fundamental - Completo ( ) Ensino Médio – Incompleto
( ) Ensino Médio – Completo ( ) Ensino Superior – Incompleto ( ) Mestrado ou Doutorado ( ) Não sei

Questões relativas ao Contexto do Curso

11. Você está respondendo o questionário sobre qual idioma?
( ) Inglês ( ) Espanhol ( ) Francês ( ) Italiano ( ) Italiano
( ) Alemão ( ) Japonês ( ) Outros:

12. Quantos anos você tinha quando concluiu o curso? _____

13. Quanto tempo faz que você finalizou o seu curso de idioma?
( ) até 1 ano ( ) de 1 a 2 anos ( ) de 2 a 3 anos ( ) 3 a 4 anos ( ) mais de 4 anos

14. Qual era a duração total do Curso?
( ) até 1 ano ( ) de 1 a 2 anos ( ) de 2 a 3 anos ( ) 3 a 4 anos ( ) mais de 4 anos

15. Sobre a frase: "Eu tenho intenção de fazer um outro curso para aprender um novo idioma".
( ) Discordo Totalmente ( ) Discordo ( ) Nem concordo, nem discordo ( ) Concordo ( ) Concordo Totalmente

16. Em relação a seguinte afirmação: "Sobre o idioma que eu já fiz o curso e finalizei, há uma possibilidade de eu voltar a fazer um outro curso sobre este mesmo idioma, afim de me aperfeiçoar"
( ) Discordo Totalmente ( ) Discordo ( ) Nem concordo, nem discordo ( ) Concordo ( ) Concordo Totalmente

17. Em uma escala de 1 a 5, qual era o seu nível de conhecimento depois da realização do curso de idioma?
( ) 1 ( ) 2 ( ) 3 ( ) 4 ( ) 5

18. Como você considerava o valor da mensalidade do curso em relação a sua renda mensal?
( ) Muito caro ( ) Caro ( ) Nem caro, nem barato ( ) Barato ( ) Insignificante

19. Assinale alternativas que relacionem situações que ocorreram durante a realização do curso (é permitido assinalar múltiplas alternativas)
( ) Separou-se/divorciou-se ( ) Começou a trabalhar ( ) Mudou-se de instituição de ensino
( ) Mudou-se de endereço ( ) Mudou o idioma estudado ( ) Perdeu o emprego
( ) Teve filho(s) ( ) Realizou viagens ( ) Enfrentou problemas de saúde
( ) Outro:_____ ( ) Nenhuma das alternativas

20. O curso era:
( ) Presencial ( ) À distância

Dificuldades dos Cursos PRESENCIAIS

(apenas para quem faz cursos presenciais)

21. Como você avaliava a distância a ser percorrida até a instituição de ensino?
( ) Muito perto ( ) Perto ( ) Nem longe, nem perto ( ) Longe ( ) Muito longe

22. Como você avaliava a facilidade de transporte até a instituição de ensino?
( ) Muito fácil ( ) Fácil ( ) Nem difícil, nem fácil ( ) Difícil ( ) Muito Difícil

23. Como você ia para o curso?
( ) Carona ( ) Bicicleta ( ) Motocicleta própria ( ) Carro próprio ( ) Transporte público

Dificuldades dos Cursos à DISTÂNCIAS
(apenas para quem faz cursos à distâncias)

21. Assinale o grau de concordância com a seguinte frase: "Eu tinha facilidade para dominar novas tecnologias, sendo assim não tive dificuldades para interagir com o ambiente virtual de aprendizagem".
( ) Discordo Totalmente ( ) Discordo ( ) Nem concordo, nem discordo ( ) Concordo ( ) Concordo Totalmente

22. Assinale o grau de concordância com a seguinte frase: "Eu não enfrentei dificuldades com o material didático nem tive problemas de comunicação com o tutor, ele sempre me estimulava e fornecia feedback sobre o meu desempenho".
( ) Discordo Totalmente ( ) Discordo ( ) Nem concordo, nem discordo ( ) Concordo ( ) Concordo Totalmente

Questões relativas ao Nível de Satisfação - Curso Presencial
(apenas para aqueles que fizeram cursos presenciais)

RESPONDA AS QUESTÕES ABAIXO, CONSIDERANDO A PERCEPÇÃO QUE VOCÊ TINHA DURANTE A REALIZAÇÃO DO CURSO, NÃO A SUA PERCEPÇÃO ATUAL.

24. Como você classificava a sua **motivação** para o estudo de idiomas em geral:
( ) Totalmente desmotivado ( ) Desmotivado ( ) Nem motivado, nem desmotivado ( ) Motivado ( )Muito Motivado

25. Qual era o seu nível de satisfação com o **idioma escolhido**:
( ) Muito insatisfeito ( ) Insatisfeito ( ) Nem satisfeito, nem insatisfeito ( ) Satisfeito ( ) Muito satisfeito

26. De acordo com sua percepção, o quanto você considerava relevante o aprendizado do idioma escolhido para a sua **carreira**?
( ) Sem importância ( ) Pouco importante ( ) Importante ( ) Muito importante ( ) Fundamental

27. Como você classificava a importância de aprender o idioma escolhido para a sua **vida pessoal**?
( ) Sem importância ( ) Pouco importante ( ) Importante ( ) Muito importante ( ) Fundamental

28. Como você avaliava seu nível de dificuldade em aprender idiomas?
( ) Muito baixa ( ) baixa ( ) Nem baixa, nem alta ( ) alta ( ) Muito alta

29. Quando era comparado o seu nível de domínio do idioma estudado com o nível dos demais alunos, você acreditava que os estudantes estavam:
( ) Muito abaixo do seu nível ( ) Abaixo do seu nível ( ) Mesmo nível ( ) Acima do seu nível ( ) Muito acima do seu nível

30. Qual era o seu grau de dificuldade para conciliar a carga horária do curso com suas demais atividades do dia a dia?
( ) Muita baixo ( ) Baixo ( ) Nem alto, nem baixo ( ) Alto ( ) Muito alto

31. Como você avaliava os horários das aulas?
( ) Péssimo ( ) Ruim ( ) Nem bom, nem ruim ( ) Bom ( ) Ótimo

32. Como era a sua integração com os demais alunos da classe?
( ) Péssimo ( ) Ruim ( ) Nem bom, nem ruim ( ) Bom ( ) Ótimo

33. Como você classificava o seu relacionamento com o professor?
( ) Péssimo ( ) Ruim ( ) Nem bom, nem ruim ( ) Bom ( ) Ótimo

34. Qual era o seu nível de satisfação com a instituição de ensino?
( ) Muito insatisfeito ( ) Insatisfeito ( ) Nem satisfeito, nem insatisfeito ( ) Satisfeito ( ) Muito satisfeito

35. Você poderia nos fornece o seu e-mail para receber os resultados obtidos e para realizarmos uma segunda etapa desta pesquisa? (Opcional)

_____

36. Caso você tenha alguma reclamação, sugestão de melhoria, dúvidas, ou feedback, por gentileza, relate-o no campo abaixo. (Opcional)

_____

Questões relativas ao Nível de Satisfação – Curso à Distância

RESPONDA AS QUESTÕES ABAIXO, CONSIDERANDO A PERCEPÇÃO QUE VOCÊ TINHA DURANTE A REALIZAÇÃO DO CURSO, NÃO A SUA PERCEPÇÃO ATUAL.

23. Como você classificava a sua **motivação** para o estudo de idiomas em geral::
( ) Totalmente desmotivado ( ) Desmotivado ( ) Nem motivado, nem desmotivado ( ) Motivado ( ) Muito Motivado

24. Qual era o seu nível de satisfação com o **idioma escolhido**:
( ) Muito insatisfeito ( ) Insatisfeito ( ) Nem satisfeito, nem insatisfeito ( ) Satisfeito ( ) Muito satisfeito

25. De acordo com sua percepção, o quanto você considerava relevante o aprendizado do idioma escolhido para a sua **carreira**?
( ) Sem importância ( ) Pouco importante ( ) Importante ( ) Muito importante ( ) Fundamental

26. Como você classificava a importância de aprender o idioma escolhido para a sua **vida pessoal**?
( ) Sem importância ( ) Pouco importante ( ) Importante ( ) Muito importante ( ) Fundamental

27. Como você avaliava seu nível de dificuldade em aprender idiomas?
( ) Muito baixa ( ) Baixa ( ) Nem baixa, nem alta ( ) Alta ( ) Muito Alta

28. Quando era comparado o nível de exigência do módulo que você estava cursando, com o seu nível de domínio do idioma estudado, você acredita que o módulo estava:
( ) Muito abaixo do seu nível ( ) Abaixo do seu nível ( ) Mesmo nível ( ) Acima do seu nível ( ) Muito acima do seu nível

29. Qual era o seu grau de dificuldade para conciliar a carga horária do curso com suas demais atividades do dia a dia?
( ) Muito baixo ( ) Baixo ( ) Nem alto, nem baixo ( ) Alto ( ) Muito alto

30. Como você avaliava a plataforma e suas funcionalidades do curso?
( ) Péssimo ( ) Ruim ( ) Nem bom, nem ruim ( ) Bom ( ) Ótimo

31. Como você avaliava a metodologia do idioma estudado?
( ) Péssimo ( ) Ruim ( ) Nem bom, nem ruim ( ) Bom ( ) Ótimo

32. Como você avalia a sua disciplina para realizar as atividades propostas no curso de idioma.
( ) Péssima ( ) Ruim ( ) Nem bom, nem ruim ( ) Bom ( ) Ótima

33. Qual o seu nível de satisfação com a instituição de ensino?
( ) Muito insatisfeito ( ) Insatisfeito ( ) Nem satisfeito, nem insatisfeito ( ) Satisfeito ( ) Muito satisfeito

34. Você poderia nos fornece o seu e-mail para receber os resultados obtidos e para realizarmos uma segunda etapa desta pesquisa? *(Opcional)*

_____

35. Caso você tenha alguma reclamação, sugestão de melhoria, dúvidas, ou feedback, por gentileza, relate-o no campo abaixo. *(Opcional)*

_____

APÊNDICE D - Questionário Caso 5

PESQUISA SOBRE O PERFIL DO ESTUDANTE DE LÍNGUAS ESTRANGEIRAS

O objetivo desta pesquisa é identificar o perfil dos indivíduos que estudam ou já estudaram línguas estrangeiras.

Tempo esperado de preenchimento: 5 minutos

TERMO DE CONSENTIMENTO LIVRE E ESCLARECIDO (TCLE)

Este questionário se trata de uma pesquisa conduzida por Monique Tamara de Lima (R.A 1838831) discente do Curso de Engenharia de Produção da Universidade Tecnológica Federal do Paraná (UTFPR) - Câmpus Londrina, sob a supervisão do Professor Doutor Rafael Henrique Palma Lima. Tendo como enfoque prever a evasão acadêmica de cursos de idiomas cujo objetivo é melhorar a experiência de aprendizado dos estudantes e reduzir a taxa de evasão, por meio da identificação de atributos e padrões que ocasionam esse evento.

Para isto estamos a pedir-lhe a sua colaboração na participação voluntária nesta pesquisa, respondendo este questionário. A sua participação é de grande importância para o sucesso desta investigação. Todavia, há a opção de não participar, ou até mesmo interromper a qualquer momento o preenchimento do questionário, sem nenhum prejuízo ou coação. Todas as informações obtidas nesse estudo são de natureza confidencial e de total anonimato dos participantes. Os resultados serão apresentados de maneira agregada publicações cientificas de cunho acadêmico.

( ) Concordo ( ) Discordo

24. Em que situação você se enquadra neste questionário?

( ) Caso 1 - Está cursando atualmente **1** curso de idioma
( ) Caso 2 - Está cursando atualmente **mais de 1** curso de idioma
( ) Caso 3 - Já fez curso de idioma, mas atualmente não está cursando
( ) Caso 4 - Já concluiu o curso de idioma, e atualmente não está cursando nenhum outro curso de línguas estrangeiras
( ) Caso 5 – Nenhum dos casos anteriores

FINALIZADO