**UNIVERSIDADE NOVE DE JULHO – UNINOVE**

**PROGRAMA DE PÓS-GRADUAÇÃO EM INFORMÁTICA E GESTÃO DO CONHECIMENTO**

**PAMELA FERREIRA ALVES ANDRELO**

**MINERAÇÃO DE DADOS EDUCACIONAIS NA IDENTIFICAÇÃO DO PERFIL DO EGRESSO DE ESCOLA TÉCNICA PÚBLICA**

Projeto de dissertação apresentado como requisito parcial da disciplina Seminários em Tecnologia da Informação Inteligente do PPGIGC da UNINOVE.

Prof. Orientador: Dr. Renato José Sassi

Ano de ingresso: 2020

Linha de pesquisa: LP3 –Tecnologia da Informação e Conhecimento

São Paulo

2020

**RESUMO** (Obrigatório)

Uma educação de qualidade permite que o jovem consiga superar dificuldades impostas pela pobreza e pela marginalização, sendo um dos primeiros obstáculos, o acesso à universidade pública ou privada. Para mudar este cenário, instituições de ensino têm buscado melhorias em suas tomadas de decisões direcionadas à captação deste jovem, por meio de levantamento de dados de seus discentes. Isto é o que ocorre com as Escolas Técnicas do Estado de São Paulo, as ETECS, que atendem cerca de 221 mil estudantes nos Ensinos Técnico, Técnico integrado ao Médio e Médio. Apesar da educação profissional técnica ser um interessante objeto de estudo, existe a necessidade de identificar o perfil de aluno, uma vez que a mesma tem seu foco no desenvolvimento educacional. Existem várias técnicas para análise de dados educacionais, sendo elas Educational Data Mining (EDM) (ou Mineração de Dados Educacionais - MDE), Learning Analytics (LA) e Academic Analytics (AA), onde se diferenciam na abordagem de cada problema. A Mineração de Dados Educacionais ou em inglês Educational Data Mining (EDM) é uma disciplina oriunda da mineração de dados, que pode auxiliar na identificação do perfil do egresso de escola técnica. A mineração de dados educacionais tem como objetivo o desenvolvimento de métodos para explorar os dados provenientes de cenários educacionais e essas metodologias são utilizadas para compreender os alunos nos seus ambientes de aprendizagem. Assim, o objetivo deste projeto é investigar a aplicação da mineração de dados educacionais na identificação do perfil de egressos da ETEC Paulistano. O estudo será desenvolvido na ETEC Paulistano, situada no Jardim Paulistano, que se encontra no centro de uma região altamente populosa e carente de escolas, de infraestrutura e de empregos. A mineração de dados computacionais será aplicada em uma base de dados, oriunda da aplicação de um questionário para 520 egressos dos anos de 2016 a 2020, que informaram o resultado dos vestibulares que participaram, considerando o curso superior aprovado, a instituição pública ou privada e no caso privada, se é bolsista ou mensalista. Espera-se que a aplicação da mineração de dados educacionais identifique o perfil do egresso da ETEC Paulistano./descrever a metodologia de pesquisa e a toda a metodologia experimental/ colocar resultados esperados

**Palavras-chave:** Mineração de Dados Educacionais, Centro Paula Souza, ETEC, Identificação do Perfil de Aluno.

(até 25 linhas)

Procure desenvolver:

- Uma frase para a contextualização da temática abordada

- Uma frase para a exposição do objetivo geral da pesquisa

- Duas ou três frases para a exposição dos autores base da plataforma teórica

- Duas ou três frases para a exposição dos métodos e materiais a serem utilizados

- Uma frase com os resultados a serem alcançados

**SUMÁRIO**

Pág.

# **1. Introdução**

Segundo a legislação educacional brasileira, o profissional técnico, é a pessoa que tem o diploma de um curso técnico atendendo as conjecturas definidas em lei: para concorrer a uma vaga em curso técnico, é necessário estar cursando ou ter concluído o ensino médio ou equivalente; os cursos técnicos devem abranger áreas especializadas e têm como objetivo atender as necessidades que vêm dos diversos setores de economia (MACHADO, 2008).

Com a Lei de Diretrizes e Bases Educacionais Nacional (LDB) de 1996, dando continuação ao que foi acordado pela Constituição Federal de 1988, é que a educação profissional técnica, em todos os seus sentidos, passa a conhecida como parte integrante e prestigiada das diferentes formas de educação previstas no País (Conselho Nacional de Educação, 2002a; 2002b).

Referindo-se à inovação tecnológica, significa uma chance única para a educação profissional e tecnológica que passa a desempenhar uma função, não exclusiva, mas indispensável para o desenvolvimento para a evolução do país. (PACHECO, 2008).

Em 1969, foi criado o Centro de Educação Tecnológica de São Paulo, denominado, a partir de 1973, Centro Estadual de Educação Tecnológica Paula Souza (CEETEPS), que atua na educação profissional do estado de São Paulo.

Mantendo cursos técnicos de nível médio em Escolas Técnicas do Estado de São Paulo, denominadas ETECS, vinculado à Secretaria de Desenvolvimento Econômico do governo do estado de São Paulo, o CPS nasceu com o propósito de expandir o ensino profissional.

As ETECS recebem mais de 220 mil estudantes nos Ensinos Técnico, Técnico integrado ao Médio e Médio, distribuídos nos 134 cursos técnicos para os setores Industrial, Agropecuário e de Serviços (CENTRO PAULA SOUZA, 2019).

As inscrições para o processo seletivo das ETECS do Centro Paula Souza são semestrais, exceto as do Ensino Médio e dos cursos técnicos integrados ao Ensino Médio que são anuais. A cada processo seletivo é preciso conferir quais os cursos oferecidos em cada unidade, pois o conselho de escola junto com a supervisão do Centro Paula Souza define quais cursos serão ofertados nas ETECs (ETEC PAULISTANO, 2019).

A Escola Técnica Estadual Paulistano (ETEC Paulistano) foi criada em 22 de dezembro de 2009 através do Plano Estadual de Expansão do Governo de São Paulo junto com o Centro Paula Souza, iniciando suas atividades no dia 22 de fevereiro de 2010 (ETEC PAULISTANO, 2019).

A ETEC Paulistano situada no Jardim Paulistano, encontra-se no centro de uma região altamente populosa, porém carente de parques, de escolas, de infraestrutura e de empregos. Nesta região não há nenhuma linha de trem ou metrô, poucas indústrias ou fábricas, sendo que o terceiro setor predomina como principal empregador.

O corpo discente da ETEC Paulistano é composto por 95% de alunos da região da Vila Brasilândia, um dos pontos mais populosos e carentes da cidade de São Paulo. Deste público, 70% deles vêm diretamente da escola pública municipal, ou seja, o perfil do aluno da ETEC Paulistano é de escola pública e de baixa renda (ETEC PAULISTANO, 2019).

Analisar o perfil dos egressos nos vestibulares da ETEC Paulistano, é importante, pois é um instrumento de gestão necessário para a melhoria da qualidade de ensino. Além do acesso à universidade por meio dos cursos de profissionalização ofertados pela ETEC Paulistano podem proporcionar aos alunos melhores condições de entrar no mercado de trabalho ou mesmo de empreender, o que possibilita independência financeira necessária para estudar e melhorar a qualidade de vida de si próprio e dos familiares.

Técnicas podem ser usadas na identificação de perfis, dentre elas, técnicas de mineração de dados, que usam algoritmos inteligentes baseados em descoberta de conhecimento por meio de procura de padrões (PIATETSKY-SHAPIRO, 1996).

A mineração de dados se caracteriza pela existência do algoritmo minerador, que perante a tarefa determinada será capaz de extrair de modo eficiente conhecimento implícito útil de uma base de dados (FAYYAD et al. (1996).

Tratando-se dados educacionais, a Mineração de Dados Educacionais ou Educational Data Mining (EDM), uma disciplina oriunda da mineração de dados, pode auxiliar na identificação do perfil do egresso de escola técnica.

Segundo o Journal of Educational Data Mining (2016), a mineração de dados educacionais tem como objetivo o desenvolvimento de métodos para explorar os dados provenientes de cenários educacionais e essas metodologias são utilizadas para compreender os alunos nos seus ambientes de aprendizagem.

Segundo Silva, 2016 a mineração de dados educacionais proporciona o conhecimento de fatores que aperfeiçoem a proposta educacional, além de prever o desempenho dos alunos e de elementos que influenciam o aprendizado.

Além da mineração de dados educacionais Silva (2017) diz que a análise de dados educacionais se divide em temáticas como Educational Data Mining (EDM) (ou Mineração de Dados Educacionais - MDE), Learning Analytics (LA) e Academic Analytics (AA), onde se diferenciam principalmente no como se trata cada tipo de problema.

Os conceitos centrais estudados neste trabalho estão correlacionados com a dificuldade que a ETEC tem para obter uma visão de seus egressos. Desta forma, ter o perfil dos egressos, poder atuar de forma eficaz e efetiva na tomada de decisão da instituição de ensino.

Apesar da educação profissional técnica ser um interessante objeto de estudo, existe a necessidade de identificar o perfil de egressos, uma vez que a mesma tem seu foco no desenvolvimento educacional.

A partir dessas considerações emerge a seguinte questão de pesquisa, que este projeto pretende responder: “**Como a mineração de dados educacionais pode auxiliar na identificação do perfil de egressos da ETEC Paulistano?**”.

No resumo, introdução e fundamentação teórica deve conter ciência de dados educacionais, MDE, LA e AA.

Introdução fraca tem que reforçar

1.1 Contextualização do tema (Estado da arte, obrigatório caso não escreva o capítulo 3) (Obrigatório)

1.2 Identificação de lacunas (possível solução) ou questão(ões) de pesquisa (Obrigatório)

1.3 Delimitação do tema (escopo e não-escopo) (Opcional)

# **2. Problema de pesquisa**

## 2.1. Objetivos

**- Objetivo geral**

O objetivo geral deste projeto é investigar a aplicação da mineração de dados educacionais na identificação do perfil de egressos da ETEC Paulistano.

Este projeto proposto está dividido em dois semestres (etapas) descritos a seguir:

* **1ª etapa** (segundo semestre de 2020);
* **2ª etapa** (primeiro semestre de 2021).
* **Objetivo da 1ª etapa** (segundo semestre de 2020)

Investigar e analisar o processo de gestão educacional e de mudanças estratégicas e suas relações com as tomadas de decisões.

**Objetivos Específicos 1ª. etapa:**

* Levantamento na literatura de questões pertinentes a visão que a gestão educacional brasileira possui do tema estratégia educacional;
* Levantamento na literatura da aplicação da Mineração de Dados Educacionais;
* Aplicação do questionário para criação da base de dados de egressos;
* Avaliação da aplicação do questionário;
* Validação das respostas dadas pelos egressos;
* Criação da base de egressos.
* **Objetivo 2ª etapa** (primeiro semestre de 2021)

Aplicação da Mineração de Dados Educacionais na identificação do perfil de egressos.

* **Objetivos específicos 2ª etapa**
* Preparação do ambiente computacional destinado à aplicação da Mineração de Dados Educacionais na base de egressos;
* Selecionar as técnicas de mineração de dados educacionais que serão aplicadas;
* Avaliar os resultados da aplicação das técnicas de mineração de dados educacionais;
* Avaliar a descoberta de conhecimento gerada pela aplicação das técnicas de mineração de dados educacionais;
* Participar do Encontro de Iniciação Científica da UNINOVE;
* Elaborar resumo e artigo que serão submetidos a congresso e periódico.

## 2.2. Justificativa para o estudo

As instituições de ensino têm buscado melhorias em suas tomadas de decisões, por meio de levantamento de dados de seus docentes e discentes. A mineração de dados educacionais tem seu foco em analisar e apoiar decisões para obter resultados para gerar estratégias para trazer mais candidatos para seu processo seletivo.

A cada processo seletivo é preciso conferir quais os cursos oferecidos em cada unidade, para que atenda melhor a comunidade em seu entorno, comunidade e empresas. O conselho de escola junto com a supervisão do Centro Paula Souza define quais cursos serão ofertados nas ETECS.

Neste contexto, a justificativa para a realização deste trabalho está pautada na necessidade de analisar o perfil dos egressos de ETEC, pois são instrumentos de gestão necessários para a melhoria da qualidade de ensino.

Uma educação de qualidade permite que o jovem consiga superar muitas dificuldades impostas pela pobreza e pela marginalização sendo, um dos primeiros obstáculos o acesso à universidade pública ou privada.

O acesso à universidade por meio dos cursos de profissionalização ofertados pela ETEC Paulistano pode oferecer condições para a entrada no mercado de trabalho, ou mesmo se tornar de empreender, o que pode permitir a independência financeira necessária para estudar ou para melhorar a sua qualidade de vida.

Isto pode ser comprovado ao se verificar que a formação nas ETECS garante emprego para 80% dos alunos e, a cada cinco pessoas que completam os cursos, quatro saem empregadas (GOVERNO DE SÃO PAULO, 2020). A empregabilidade de alunos formados está em torno de 92% (CENTRO PAULA SOUZA. 2020).

Os conceitos centrais estudados neste trabalho estão correlacionados com a dificuldade que a ETEC tem para obter uma visão de seus egressos. Desta forma, ter o perfil dos egressos, poder atua de forma eficaz e efetiva na tomada de decisão da instituição de ensino.

## 2.3. Delimitação do tema

Segundo Lakatos e Marconi (2010), os limites de uma pesquisa podem ser determinados por meio de três tópicos: assunto, extensão e pela série de fatores. Afirmam ainda que essa limitação pode ser determinada quanto ao objeto de estudo, campo de investigação e nível de investigação.

Quanto ao objeto de estudo, este trabalho delimita-se em colher informações referentes a Identificação de Perfil e Mineração de Dados Educacionais dos egressos da ETEC Paulistano.

Já quanto ao campo de investigação, a pesquisa será expandida para os temas que tem correlação com esses assuntos, como Identificação do Perfil do Aluno da ETEC Paulistano, Dados Educacionais e Mineração de Dados Educacionais.

O nível de investigação será limitado a projeto de Mineração de Dados Educacionais na identificação de perfil dos egressos da ETEC Paulistano, já que este será o ambiente onde a pesquisa será aplicada.

A ETEC Paulistano tem a necessidade de gerar estratégias afim de trazer mais candidatos para o seu processo seletivo. Para isso será importante analisar o desempenho dos egressos nos vestibulares de Universidades Públicas e Privadas.

2.1 Situação problema (conceitos centrais estudados e o problema de pesquisa) (Opcional)

2.2 Justificativa da pesquisa (relevância, importância e atualidade) (Obrigatório)

2.3 Proposições / hipóteses de resolução do problema (Opcional)

2.4 Objetivos geral e específicos (Obrigatório)

2.5 Modelo teórico (constructo teórico previsto) (Opcional)

2.6 Delimitação da pesquisa (Opcional)

**3. Plataforma teórica prevista** (Opcional)

## [**3.1. Gestão Educacional Brasileira**](#_Toc5230403)

As tecnologias digitais, a reforma do mundo físico, as modificações do ser humano e a integração disso tudo com o ambiente, são os pilares das transformações que estamos começando a vivenciar na gestão educacional brasileira.

Essa realidade se mostra complexa para os trabalhadores da área social, em particular os da educação e sua administração, tanto no âmbito estatal como na iniciativa privada.

Sander (2005) expõe que o termo administração dominou a tempos, o pensar e o fazer a educação. No entanto, uma série de termos estão em disputa de espaço, destacando-se os de gestão e gerência. No entanto, o termo gestão era praticamente inexistente na teoria e na prática da educação brasileira e, os que primeiro o utilizaram, eram encarados com desconfiança e desaprovação, já que os primeiros a adotar este termo eram os administradores de empresa.

De fato, o uso do termo gestão vem aumentando no segmento administrativo, tanto no setor público em geral, como na educação, especificamente. É inclusive respaldado pela constituição de 1988, pela flexibilização oferecida pela Lei de Diretrizes e Bases da Educação Nacional (LDB) de 1996 e por diversos instrumentos normativos (CORDAS, 2016).

Independente da origem etimológica e do conceito, a gestão da educação no Brasil se equipara à administração da educação. Nesse sentido, a gestão da educação abarca desde a formulação de políticas e planos institucionais e a concepção de projetos pedagógicos para os sistemas educacionais e as instituições escolares até a execução, a supervisão e a avaliação institucional das atividades de ensino, pesquisa e extensão e a administração dos recursos financeiros, materiais e tecnológicos (SANDER, 2005).

## 3.2. Escolas Técnicas Estaduais

Segundo o sítio do Centro Paula Souza as Escolas Técnicas Estaduais (ETECS) são caracterizadas por: 223 ETECS, distribuídas por 165 municípios paulistas. As ETECS atendem mais de 208 mil estudantes nos Ensinos Técnico, 151 cursos, voltados a todos os setores produtivos públicos e privados (Centro Paula Souza, 2019).

A administração colaborativa é um dos elementos que guiam a gestão das ETECS, estando predita no Regimento Comum das Escolas Técnicas Deliberação CEETEPS Nº 003, de 18-7-2013 (Brasil, 2013). Essa determinação estabelece conceitos de administração igualitária guiarão a gestão da ETEC, respeitando as relações fundamentadas na comunicação e na conformidade, trazendo como prática a cooperação, o debate coletivo e a independência.

A cooperação possibilitará que todo corpo acadêmico envolvido no processo de tomada de decisões para a instituição e para o desenvolvimento da ETEC, prospéra um ambiente de trabalho oportuno com uma harmonização entre todas as divisões das ETECS.

Segundo o Plano Plurianual de Gestão (PPG) de 2016 a ETEC Paulistano situada no Jardim Paulistano, se encontra no centro de uma região altamente populosa, porém também uma região carente, carente de parques, de escolas, de infraestrutura e de empregos. Nesta região da escola não há nenhuma linha de trem ou metrô, poucas industrias ou fábricas, sendo que o terceiro setor predomina como principal empregador, tirando as situações de subemprego e de comércio familiar.

A ETEC Paulistano situada então nessa região, atende a todos moradores do entorno deste ambiente com alunos moradores do Jardim Paulistano e entorno, Jardim Carumbé, Vila Zatti, Jardim Nossa Senhora Aparecida, Morro Grande, Jardim Paquetá e COAB Vila Nova Brasilândia. Uma região com cerca de 300.000 habitantes, percebe-se aí que o distrito da Brasilândia é um dos mais populosos da cidade de São Paulo.

O corpo discente da escola é composto por 95% de alunos da região Brasilândia e Pirituba, sendo que alguns veem de regiões ainda mais afastadas, como Parada de Taipas e Perus. Deste público 70% deles veem diretamente da escola pública municipal. O que demonstra claramente o perfil do aluno da ETEC Paulistano que é um aluno de escola pública e de baixa renda.

Uma educação de qualidade permite que o jovem consiga superar muitas dificuldades impostas pela pobreza e pela marginalização, um dos primeiros obstáculos é o acesso à universidade pública e ou gratuita.

Além do acesso à universidade por meio dos cursos de profissionalização ofertados pela ETEC muitos jovens têm mais condições de entrar no mercado de trabalho ou mesmo de empreender, o que permite a independência financeira necessária para estudar ou para melhorar a qualidade de vida de si próprio ou dos familiares.

Diante de todo este cenário, a ETEC Paulistano tem por missão ser referência de um local de estudo, profissionalização, mas também de cultura e de lazer para a população da região para que possa transformar a vida do discente bem como das famílias. E a escola tem feito isso por meio da profissionalização que auxilia a muitos a conseguirem o primeiro emprego, bem como ao ingresso nas universidades públicas e privadas por grande parte dos alunos, estando a ETEC Paulistano muito acima dos índices de acesso à universidade propostos pelo Índice de Desenvolvimento da Educação Básica (IDEB) (ETEC PAULISTANO, 2020).

Além disso, a ETEC Paulistano figura-se com a posição 27º entre as melhores escolas públicas de São Paulo, e com a posição 3220 entre as 27780 escolas analisadas pelo (IDEB) (ETEC PAULISTANO, 2020).

## 3.3. Dados Educacionais

## Os dados são valores referentes a medições, contagens ou observações relativas a uma amostra ou a um determinado fenômeno. Como exemplo de dados, podemos citar a nota de um aluno em uma avaliação, a quantidade de acessos de um aluno a um material, e a presença ou ausência de um aluno em uma aula (SILVA, 2016).

## No contexto educacional, os dados podem ser de diferentes fontes como, por exemplo, ambientes virtuais de aprendizagem, questionários, sites de professores, sistemas acadêmicos, sistemas tutores inteligentes, dentre outros, e fornecer informações sobre estudantes, professores e os contextos educacionais em que estão inseridos (NASCIMENTO et al., 2016).

## De acordo com Romero, Romero e Ventura (2014), os dados educacionais têm alguns atributos específicos e decorrentes da condição singular de onde são retirados. Por exemplo, muitas vezes os alunos deixam de fazer os exercícios que lhe são propostos de uma aula qualquer, gerando informações faltantes na base de dados, deixando campos vazios ou imcompletos.

## Conforme Gandomi e Haider (2015) a diversidade das informações é referente à diferença basilar da base de dados, podendo utilizar-se de modelos de dados estruturados, semiestruturados e não estruturados, dado que hoje em dia grande parte dos dados vem de de textos, imagens, áudios, vídeos ou sensores, que são exemplos de dados não estruturados.

## As informações sobre anos de estudo, frequência de nível escolar e frequência na rede pública ou privada, classificadas por sexo e por grupos de idade são importantes indicadores, que podem ser analisados para Brasil ou para grandes regiões. Pode-se, a partir deles calcular indicadores, como a taxa de analfabetismo, a taxa de escolarização bruta e líquida, a taxa de frequência à escola e a proporção da população com anos de estudo. (RIGOTTI e CERQUEIRA, 2015).

## 3.4. Análise de dados educacionais

A área educacional vem integrando novas tecnologias por conta dos diferentes tipos de abordagens educacionais produzirem cada vez mais dados, sendo assim, demandam análises minuciosas para que tenha um melhor planejamento e execução de ações no processo educativo (Silva et al., 2017).

A análise de dados educacionais tem o intuito de entender os estudantes e o contexto em que aprendem, dessa maneira, retrata uma área de pesquisa em ascensão em Informática e Educação para desenvolvimento de técnicas e métodos que investigam os dados gerados em ambientes educacionais e de gestão educacional (Daniel, 2016).

Para Silva et al (2017), a análise de dados educacionais, se divide em temas, sendo eles Academic Analytics (AA), Learning Analytics (LA) e Educational Data Mining (EDM) (ou Mineração de Dados Educacionais - MDE).

O objetivo da Academic Analytics ou AA, é procurar entender os dados de registro dos alunos, dados estes, que vem de sistemas de informação da própria Instituição de Ensino, e relacionar com a vivência acadêmica dos alunos na Instituição (Campbell e Oblinger, 2007).

O termo Learning Analytics, de acordo com Souza et al (2016), é um processo para intermediação, coleta, análise e relação dos dados de alunos e seu ambiente de aprendizagem, com o propósito de otimizar e compreender o ensino-aprendizagem em que esse processo acontece (Siemens et. al., 2011).

Por outro lado a Educational Data Mining (EDM) (ou Mineração de Dados Educacionais - MDE) tem a finalidade de fazer a descoberta de conhecimento através do comportamento dos estudantes e o cenário onde aprendem, viabilizando elementos para que o professor, gestor educacional ou até mesmo o aluno avalie eventuais padrões descobertos (Romero e Ventura, 2007; Romero et al., 2016; Ducange et al., 2016).

## 3.5. Mineração de Dados

### A mineração de dados se caracteriza pela existência do algoritmo minerador, que perante a tarefa determinada será capaz de extrair de modo eficiente conhecimento implícito útil de uma base de dados (FAYYAD et al., 1996).

### Os padrões extraídos devem ser confiáveis, compreensíveis e úteis, empregando conhecimento com utilidade e com proveito de alguma vantagem, seja científica ou comercial, como por exemplo, identifica padrões e descobrir informações relevantes que auxiliam na tomada de decisão, citando o caso da escolha das disciplinas de um plano de curso, cursos ofertados pela escola, entre outras (DANTAS et al., 2008).

### De acordo com FAYYAD et al. (1996), o processo de KDD é formado pormuitas etapas, pode ser considerado como o processo de descoberta de padrões e tendências por estudo de amplos coleções de dados, sendo o processo de mineração de dados a etapa principal, tratando-se da execução prática de análise e de algoritmos específicos que, sob insuficiências computacionais atendíveis, gera uma conexão especifica de padrões a partir de dados.

### **3.5.1. Fases do KDD**

Descreve- se a seguir as fases do KDD segundo FAYYAD et al. (1996):

Após a definição do objetivo a ser alcançado, o foco passa a ser a escolha ou seleção dos dados a serem minerados, podendo ser um conjunto de dados ou um subconjunto de variáveis onde a extração será realizada.

A fase de Pré-Processamento tem por objetivo assegurar a qualidade dos dados envolvidos no KDD realizando operações básicas como a remoção de ruídos, que podem ser, por exemplo, atributos nulos.

A fase seguinte, formatação consiste na seleção e transformação dos dados em que serão selecionados os atributos realmente interessantes ao usuário, além de transformados utilizando o padrão ideal para aplicar algoritmos de mineração.

Após a realização das fases anteriores, a Mineração de Dados (Data Mining) é iniciada. Esta fase é a mais importante do KDD, sendo realizada através da escolha do método e do algoritmo mais compatível com o objetivo da extração, a fim de encontrar padrões nos dados que sirva de subsídios para descobrir conhecimentos ocultos.

A Interpretação e Avaliação ou Pós-Processamento é a fase que identifica, entre os padrões extraídos na etapa de Data Mining, os padrões interessantes ao critério estabelecido pelo usuário, podendo voltar à fase inicial para novas iterações.

Ao término da avaliação, o conhecimento descoberto poderá ser implantado e incorporado ao sistema, sempre documentando e publicando os métodos, a fim de apresentar o conhecimento descoberto ao usuário.

Apresenta-se na Figura 1 as fases do KDD (FAYYAD et al. 1996).

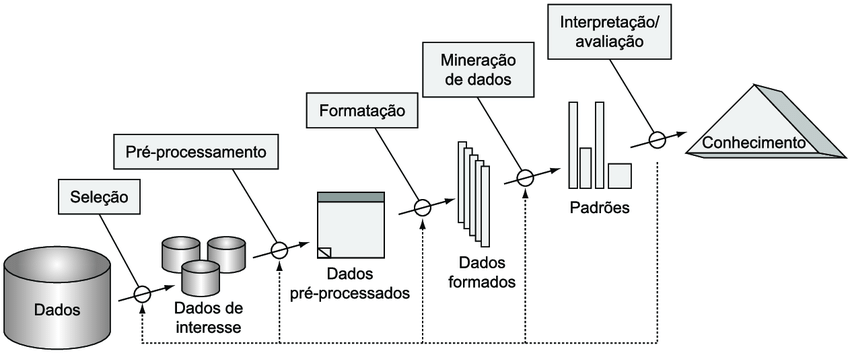


Figura 1. Fases do Processo de KDD. Fonte: Fayyad et al. (1996).

Dentre as fases descritas na Figura1, destaca-se a mineração de dados, pois é considerada a etapa mais importante do processo de KDD. Descrita pela presença do algoritmo minerador, que será capaz extrair de modo eficiente conhecimento implícito e útil de um banco de dados, através da tarefa determinada (FAYYAD, et al., 1996).

Mineração de Dados é o processo de pesquisa em bases de dados para extração de conhecimento, utilizando técnicas estatísticas ou de Inteligência Artificial para procurar relações de similaridade ou discordância entre dados, com o objetivo de encontrar padrões, irregularidades e regras, com o intuito de transformar dados, aparentemente ocultos, em informações relevantes para a tomada de decisão e/ou avaliação de resultados (AGRAWAL et al., 1993).

De acordo com Fayyad et al. (1996), existem tarefas para encontrar respostas ou extrair conhecimento em bases de dados, aplicando mineração de dados, sendo as mais comuns: Regras de Associação ou Análise de Afinidade, Classificação, Agrupamento.

Apresenta-se a seguir um resumo das tarefas mais comuns (FAYYAD, et al., 1996):

* As Regras de Associação ou Análise de Afinidade, também chamada de Análise de Cestas de Mercado (*Market Basket Analysis*). A ideia principal é identificar grupos de itens tipicamente associados (o que vai com o quê?). O objetivo é encontrar tendências, a partir de grande número de transações que possam ser usadas para entender e explorar padrões de compra.
* Classificação, cria um modelo automático, a partir de um conjunto inicial de registros que servem de exemplo e é chamado de conjunto de treinamento. Estes registros devem pertencer a um grupo de classes pré-definidas. O modelo é formado de padrões, essencialmente generalizações em relação aos registros, usados para diferenciar as classes. Uma vez obtido o modelo, ele será usado para classificar automaticamente os demais registros.
* Agrupamento ou Clusterização, agrupa os registros em conjuntos comparativamente menores, segmentos. Esse agrupamento é realizado por meio da identificação de características similares que distinguem o conjunto de dados. O objetivo nessa tarefa é de maximizar similaridade intra-cluster e minimizar a similaridade inter-cluster.

### **3.5.2.** [**Mineração de Dados Educacionais**](#_Toc5230405)

A mineração de dados se caracteriza pela existência do algoritmo minerador, que perante a tarefa determinada será capaz de extrair de modo eficiente conhecimento implícito útil de uma base de dados (FAYYAD et al., 1996).

Os padrões extraídos devem ser confiáveis, compreensíveis e úteis, empregando conhecimento com utilidade e com proveito de alguma vantagem, seja científica ou comercial, como por exemplo, identifica padrões e descobrir informações relevantes que auxiliam na tomada de decisão, citando o caso da escolha das disciplinas de um plano de curso, cursos ofertados pela escola, entre outras (DANTAS et al., 2008).

De acordo com FAYYAD et al. (1996), o processo de KDD é formado pormuitas etapas, pode ser considerado como o processo de descoberta de padrões e tendências por estudo de amplos coleções de dados, sendo o processo de mineração de dados a etapa principal, tratando-se da execução prática de análise e de algoritmos específicos que, sob insuficiências computacionais atendíveis, gera uma conexão especifica de padrões a partir de dados.

3.1 Principais tópicos a serem desenvolvidos

3.2 Principais autores (seminais, estado da arte e autor-base de cada tópico)

Colocar uma tabela especificando o assunto e o principal autor

Pegar a tabela da tese do domingos (tese ppgi)

**4. Métodos e instrumentos de pesquisa previstos**

# 4 Materiais e Métodos

## 4.1. Caracterização Metodológica

A metodologia de pesquisa adotada neste trabalho foi definida, a partir do ponto de vista de sua natureza, como pesquisa aplicada, visto que tem como objetivo gerar conhecimento para a solução de problemas, possuindo assim uma aplicação prática (GIL, 2002).

Pode ser considerada também como pesquisa exploratória que visa proporcionar maior familiaridade com o problema com vistas a torná-lo explícito ou a construir hipóteses. Do ponto de vista de sua abordagem é de natureza qualitativa, cujo ambiente de pesquisa teve como fonte direta os dados levantados (YIN, 2016).

A pesquisa experimental também está presente, pois determina um objeto de estudo, selecionam-se as variáveis que seriam capazes de influenciá-lo definem-se as formas de controle e de observação dos efeitos que a variável produz no objeto (YIN, 2016).

Um questionário será aplicado aos 520 egressos da ETEC Paulistano. Assim, segundo Gil (2002), pode-se definir questionário como a técnica de investigação composta por um conjunto de questões que são submetidas a pessoas com o propósito de obter informações sobre conhecimentos, crenças, sentimentos, valores, interesses, expectativas, aspirações, temores, comportamento presente ou passado.

A realização da pesquisa bibliográfica será embasada em consultas a fontes bibliográficas e de referencial teórico: artigos, livros, teses, dissertações, websites com conteúdo sobre Mineração de Dados Educacionais, Gestão Escolar Brasileira, Identificação de Perfil.

## 4.2. Materiais e Métodos

O universo da pesquisa abrangerá uma das Escolas Técnicas do Centro Paula Souza, a ETEC Paulistano que servirá de base para esta pesquisa, bem como suas áreas, processos, docentes e discentes, que serão impactados com a implantação do projeto de mineração de dados educacionais.

A mineração de dados computacionais será aplicada em uma base de dados da ETEC Paulistano, oriunda da aplicação de um questionário para 520 egressos dos anos de 2016 a 2020, que informaram o resultado dos vestibulares que participaram, considerando o curso superior aprovado, a instituição pública ou privada, caso privada e se é bolsista ou mensalista.

A amostra da pesquisa será uma base representativa dentro do universo dos alunos egressos que prestarão vestibulares que deverão responder à pesquisa. Importante destacar que egressos que não responderem, não serão consideradas na amostragem.

A coleta de dados se dará, a partir da criação da base de dados, formada pela entrada de dados, levantamento de dados, questionário e pré-processamento. Será aplicado um questionário no Google Form, disponibilizado na rede social Facebook, solicitando que os egressos respondam as cinco questões a seguir:

A - Você é egresso de qual curso da ETEC Paulistano?

B - Qual o resultado do vestibular que você participou?

C - No caso de aprovação, qual a área do curso superior?

D - Qual o tipo de instituição: pública ou privada?

E - No caso de instituição privada, obteve bolsa de estudos parcial ou integral?

A plataforma de hardware que será utilizada nos experimentos é um computador com processador *Intel® Core ™i5* de 2,70 GHz com 4,00 GB de memória RAM DDR3, 500 GB de disco rígido e sistema operacional *Windows* 10 de 64 *bits*.

Para mapear os processos e realizar os experimentos, será utilizado o software WEKA (http://www.cs.waikato.ac.nz/ml/weka/). Desenvolvida na linguagem Java pela Universidade de Waikato na Nova Zelândia, trabalha com diversas técnicas de Data Mining, além de ser um software livre e de fácil manuseio.

## 

## 4.3. Condução dos experimentos

Os experimentos computacionais serão realizados em quatro etapas descrita na Figura 2.

**Etapa 1 :**

2019  
2020

Figura 2 As 4 etapas de realização dos experimentos computacionais. Fonte: Autor (2020)

2016  
2017  
2018

**Etapa 4 :**

Descoberta de Conhecimento

**Etapa 5 :**

Implementação do processo

**Etapa 6 :**

Controle e Manutenção

**Etapa 7 :**

Refinamento

Interpretação

e Avaliação

**Etapa 3 :**

**Etapa 2 :**

Mineração  
de Dados Educacionais

Pré-processamento

**Etapa 1 :**

**Etapa 1 :**

Questionário  
para Egressos

Levantamento de Dados

**Etapa 1 :**

Criação da Base de Dados

Descreve-se a seguir as sete etapas de realização dos experimentos computacionais:

* **Etapa 1:** Criação da base de dados de egressos formada pela entrada de dados, levantamento de dados, questionário e pré-processamento. Será solicitado ao egresso que responda a um questionário no Google Form, disponibilizado na rede social Facebook as cinco questões descritas acima.
* **Etapa 2:** Mineração de dados educacionais: Será utilizado o software WEKA para mapear os processos e realizar o experimento.
* **Etapa 3:** Interpretação e avaliação dos resultados – Será realizada por meio da identificação dos perfis descobertos no processo de mineração de dados educacionais.
* **Etapa 4:** Descoberta de conhecimento - Ao analisar os resultados, a ETEC Paulistano decidirá tomar decisões, a partir do conhecimento obtido.
* **Etapa 5:** Implementação do processo – Implementar o processo para as próximas turmas.
* **Etapa 6:** Controle e manutenção – Controlar e fazer a manutenção do processo para novos cursos e turmas.
* **Etapa 7:** Refinamento - o ajuste fino dos processos visando a melhoria do desempenho.

# **4.4. Pesquisa Bibliográfica**

**REFORÇAR MUITO**

**Está fraca**

**Aprofundar a análise --- VosViwer**

Foi realizada uma pesquisa bibliográfica sobre os temas centrais abordados na pesquisa: Centro Paula Souza, ETEC, Perfil de Aluno, Mineração de Dados Educacionais. O critério utilizado para seleção dos artigos foi pautado na relevância, atualidade, conteúdo e abrangência dos temas abordados nas publicações. Foram selecionados 38 artigos, conforme Tabela 1.

Tabela 1: Número de Publicações por Base de dados pesquisadas.

|  |  |
| --- | --- |
| **Bases de Dados Pesquisadas** | **Número de Publicações** |
| Google Academic | 13 |
| IEEExplore | 07 |
| Scielo | 11 |
| Scient Direct | 08 |
| **Total** | **38** |

Fonte: Autor (2020)

A análise dos resultados permitiu perceber a concentração de publicações sobre os temas centrais de forma correlacionada na base do Google Acadêmico. Em seguida, realizou-se uma análise temporal, para desta forma, identificar a periodicidade das publicações, os resultados temporais das publicações estão relacionados na Tabela 2.

Tabela 2: Resultados temporais das publicações.

|  |  |  |
| --- | --- | --- |
| **Análise Temporal das Publicações** | | |
| **Ano da publicação** | **Quantidade de publicações** | **Percentual de publicações** |
| 2019 | 07 | 18,42% |
| 2018 | 05 | 13,16% |
| 2016 | 04 | 10,53% |
| 2014 | 01 | 2,63% |
| 2013 | 01 | 2,63% |
| 2010 | 01 | 2,63% |
| 2009 | 03 | 7,89% |
| 2008 | 06 | 15,79% |
| 2007 | 01 | 2,63% |
| 2006 | 01 | 2,63% |
| 2005 | 04 | 10,53% |
| 2003 | 02 | 5,26% |
| 2002 | 02 | 5,26% |
| 1996 | 01 | 2,63% |

Tabela 1 - Fonte: Autor (2020)

Verificou-se que a concentração de publicações sobre os temas centrais ocorreu principalmente no período de 2016 a 2019, sendo que em 2019 se encontrou a maior quantidade de publicações 17,94%. Pode-se verificar que ao somar as publicações dos anos de 2016, 2018 e 2019 chega-se a 16 publicações em um total de 38 publicações, o que pode indicar o maior interesse pelos temas centrais da pesquisa.

Em seguida foi realizada a classificação geográfica para identificar onde se encontra a concentração das publicações, conforme apresentado na Tabela 3.

Tabela 3: Classificação Geográfica das Publicações.

|  |  |
| --- | --- |
| **País de origem da publicação** | **Quantidade de publicações** |
| Brasil | 26 |
| Canadá | 02 |
| Reino Unido | 04 |
| Estados Unidos da América | 06 |

Tabela 2 - Fonte: Autor (2020)

É notório que desde 1996 a tendência é de aumento na quantidade de publicações sobre o tema Mineração de Dados. Ressalta-se que nessa escala foram consideradas as datas das publicações que abordam um ou dois dos temas, pois se separarmos a Mineração de Dados Educacionais, poderemos ter uma distorção nessa escala, uma vez que o tema vem sido explorado há menos tempo.

Nota-se uma tendência de acréscimo na quantidade de publicações a partir do ano de 2016 e a mesma vem se mantendo ao longo dos anos, tendo o seu ápice no ano de 2019.

Evidenciou-se que a maior concentração das publicações é no Brasil. Os temas centrais vêm sendo usados para tomada de decisões em instituições educacionais. Os Estados Unidos da América se tornaram a segunda região com maior concentração. No Brasil foi possível identificar a quantidade de 26 publicações, o que pode ser considerado o mais relevante se comparado com os demais continentes. Também foi realizada a classificação das publicações por tipo de pesquisa, conforme apresentado na Tabela 4.

Tabela 4: Classificação das Publicações por Tipo de Pesquisa

|  |  |
| --- | --- |
| **Classificação das publicações por tipo de pesquisa** | |
| **Classificação das publicações** | **Quantidade de publicações** |
| Bibliográfica | 19 |
| Estudo de caso | 2 |
| Estudo exploratório | 3 |
| Experimental | 14 |

Tabela 3 - Fonte: Autor (2020)

Essa classificação teve como objetivo identificar como os temas tem sido tratado nas publicações e nota-se que a maioria das pesquisas tem embasamento bibliográfico, seguido pelo experimental. Nota-se poucos casos de estudo de dado e exploratório, apenas 5 publicações tratam se classificam nessa categoria.

Os resultados iniciais obtidos, após a pesquisa bibliográfica, demonstraram que não foram encontrados até o momento materiais que correlacionem os temas centrais que serão base do projeto a ser desenvolvido baseado em Mineração de Dados Educacionais na identificação do perfil de egressos de escolas técnicas.

## 4.5. Limitações da pesquisa

Nota-se que o tema abordado, identificação de perfil utilizando mineração de dados educacionais, pode ser mais aprofundado, principalmente no que diz respeito as aplicações desses dados na tomada de decisão.

Entretanto como o foco não está relacionado com tomada de decisão e essa pesquisa está sendo desenvolvida em um âmbito de identificar o perfil de egressos, não se pretende sair dessa área de interesse, ou seja, identificação de perfil de com mineração de dados educacionais.

4.1 Escolha e justificativa da tipologia da pesquisa (Opcional)

(natureza, abordagem, objetivo e método de pesquisa)

4.2 Universo, amostragem, amostra (Opcional)

4.3 Instrumentos de pesquisa previstos (Obrigatório)

4.4 Técnicas de coleta e tratamento de dados previstas (Opcional)

4.5 Limitações da pesquisa (Obrigatório) – Oq fiz e restrições ?

4.6 Modelo teórico-empírico previsto (constructo teórico-aplicado previsto) (Opcional)

**5. Resultados Esperados OU Resultados Preliminares** (Obrigatório)

Espera-se que técnicas baseadas em Mineração de Dados Educacionais identifiquem o perfil de egressos da ETEC Paulistano e com base nessa identificação definir estratégias de captação de alunos para os cursos oferecidos.

- Expor a estrutura de tópicos sugerida para este capítulo

Deverão constar os resultados parciais da pesquisa (caso o aluno já os tenha) ou os resultados esperados, devidamente justificados e discutidos em face ao estado da arte da área de pesquisa elegida.

**6. Trabalhos futuros Previstos** (Obrigatório) -

Apresentar as etapas e metas a serem desenvolvidas no decorrer do projeto, bem como um cronograma de atividades considerando o tempo disponível até a finalizado do curso.

**7. Considerações Finais** (Obrigatório)

Apresentar a síntese pessoal do aluno, de modo a expressar sua compreensão sobre o assunto que é objeto deste projeto bem suas repercussões no aporte ao conhecimento teórico e prático oriundos desta pesquisa.

**8. Contribuição para a área** (Obrigatório)

Contribuição para academia (algo q ninguém fez e ta crescendo.... na minha revisão sistemática ....)

Sociedade – vai ajudar ..... melhoria de desempenho

Empresa – Como vai ajudar o CPS

Apresentar os potenciais produtos (artigos/ congressos/ trabalhos técnicos) a serem gerados por este projeto, bem como a estratégia de divulgação junto à comunidade científica internacional

**9. Referências** (Obrigatório)

AGRAWAL, R; IMIELINSKI, T; SWAMI, A. **Mining association rules between sets of items in large databases**. Proc. of the ACM SIGMOD, p. 207–216, 1993.

BAKER, R.S.J.D., DE CARVALHO, A.M.J.A., RASPAT, J., ALEVEN, V., CORBETT, A.T., KOEDINGER, K.R**. Educational Software Features that Encourage and Discourage "Gaming the System". In Proceedings of the International Conference on Artificial Intelligence in Education**, pagínas 475- 482. 2009.

BRANDÃO, M. F. R., SANTOS RAMOS, C. R., AND TRÓCCOLI, B. T. (2003). **Análise de agrupamento de escolas e núcleos de tecnologia educacional: mineração na base de dados de avaliação do programa nacional de informática na educação**. In Anais do Simpósio Brasileiro de Informática na Educação (SBIE), volume 1, pages 366–374.

BRASIL, **Deliberação CEETEPS Nº 003, de 18-7-2013. Aprova o Regimento Comum das Escolas Técnicas Estaduais do Centro Estadual de Educação Tecnológica Paula Souza, publicado no DOE em 28/08/2013**. Acesso em 08/05/2018. Disponível em: http://www.centropaulasouza.sp.gov.br/etec/regimento-comum/default.asp

CAMPOS, Rodrigo Vieira. **Implicações da avaliação institucional na gestão de uma escola técnica do Centro Paula Souza.** Brasil, 2018. Disponível em: <http://www.portal.cps.sp.gov.br/pos-graduacao/trabalhos-academicos/dissertacoes/gestao-e-desenvolvimento-da-educacao-profissional/2018/rodrigo-vieira-campos.pdf>. Acesso em: 9 mar. 2020.

CECHINEL, Cristian; CAMARGO, S. D. S. **Mineração de dados educacionais: avaliação e interpretação de modelos de classificação**. Brasil, 2018. Disponível em: <https://metodologia.ceie-br.org/wp-content/uploads/2018/10/cap2_12.pdf>. Acesso em: 22 fev. 2020.

CENTRO PAULA SOUZA. **Sobre o Centro Paula Souza**. https://www.cps.sp.gov.br/sobre-o-centro-paula-souza/. Acesso em: 10 mar. 2020.

CENTRO PAULA SOUZA. Empregabilidade de alunos formados pelas Fatecs é de 92%. Disponível em: https://www.cps.sp.gov.br/empregabilidade-de-alunos-formados-pelas-fatecs-e-de-92/. Acesso em: 05 de abr. 2020.

CHASSIGNOL, M. et al. **Artificial Intelligence trends in education: a narrative overview. Procedia Computer Science**, v. 132, p. 834-842, dez./2008. Disponível em: https://www.sciencedirect.com/science/article/pii/S1877050918308275. Acesso em: 29 ago. 2019.

CONSELHO NACIONAL DE EDUCAÇÃO. Parecer CNE/CP nº 29, de 12 de dezembro de 2002. Trata das Diretrizes Curriculares Nacionais no nível de tecnólogo. **Diário Oficial da República Federativa do Brasil**, Brasília, DF, 13 dez. 2002ª.

CORDAS, V. F. C; **Fundamentos e Práticas em Educação Profissional e Tecnológica: Maturidade Vocacional e Formação Tecnológica**. 04. ed. São Paulo: CEETEPS, 2016. p. 56-57.

COSTA, L. D. S. D. et al. **IMPORTÂNCIA E NECESSIDADE DE FORMAS DE ORGANIZAÇÃO E GESTÃO ESCOLAR**. Brasil, 2018. Disponível em: <http://revistafacesa.senaaires.com.br/index.php/revisa/article/view/325/236>. Acesso em: 9 mar. 2020.

Daniel, B. K. (Ed.). (2016). **Big Data and Learning Analytics in Higher Education:**

**Current Theory and Practice**. Springer.

DANTAS, E. **O Uso da Descoberta de Conhecimento em Base de Dados para Apoiar a Tomada de Decisões**. João Pessoa, 2008. Disponível em: <https://www.aedb.br/seget/arquivos/artigos08/331_331_Artigo_SEGET_EJDR_Versao_Final_010808.pdf> Acesso em: 19/03/2020.

DIGIAMPIETRI, L. A., NAKANO, F., AND DE SOUZA LAURETTO, M. (2016). **Mineração de dados para identificação de alunos com alto risco de evasão: Um estudo de caso.** Revista de Graduação USP, 1(1):17–23.

ETEC PAULISTANO. **A Etec Paulistano**. Disponível em: http://www.etecpaulistano.com.br/?page\_id=55. Acesso em: 2 abr. 2020.

ETEC PAULISTANO. **PPG**. Disponível em: http://www.etecpaulistano.com.br/wp-content/uploads/2015/05/ppg\_2016\_etecpaulistano.pdf. Acesso em: 22 out. 2019.

FAYYAD, U.,PIATESKY-SHAPIRO, G., SMYTH, P. **Knowledge Discovery and Data Mining: : Towards a Unifying Framework**. USA, 1996. Disponível em: https://www.aaai.org/Papers/KDD/1996/KDD96-014.pdf. Acesso em: 19 mar. 2020.

GANDOMI, A.; HAIDER, M. **Beyond the hype: bid data concepts, moods, and analytics.** International Journal of Information Management, [S.l.], v. 35, n. 2, p. 137-144, 2015.

GIL, A. C. **Métodos e Técnicas de Pesquisa Social. 6ª Edição,** São Paulo: Atlas, 2002.

GOVERNO DE SÃO PAULO. **Formação na Etec garante emprego para 80% dos alunos**. Disponível em: <https://www.saopaulo.sp.gov.br/ultimas-noticias/formacao-na-etec-garante-emprego-para-80-dos-alunos/>. Acesso em: 05 de abr. 2020.

KOCHE, J. C. **Fundamentos de Metodologia Científica.** Petrópolis: Vozes, 2003.

IDEB. **IDEB Resultados e Metas**. Disponível em: http://ideb.inep.gov.br/resultado/. Acesso em: 4 abr. 2020.

JOURNAL OF EDUCATIONAL DATA MINING. **Journal of Educational Data Mining**. [S.l.], 2016. Disponível em: http://www.educationaldatamining.org/JEDM . Acesso em: 26/01/2016.

MACHADO, Lucília**. O professional tecnólogo e sua formação**. Revista da RET, ano II, n.3, p. 1-28, 2008.

MARIA, W., DAMIANI, J. L., AND PEREIRA, M. (2016). **Rede Bayesiana para Previsão de Evasão Escolar. In Anais dos Workshops do Congresso Brasileiro de Informática na Educação**, volume 5, pages 920–929.

MEATO, L. D. S; FONTANA, V. S.; CARLETTI, E. Z. B**. MINERAÇÃO DE DADOS APLICADA NA BASE DO CADASTRO ÚNICO PARA PROGRAMAS SOCIAIS COM ENFOQUE NOS**  **PERFIS DOS USUÁRIOS CADASTRADOS UTILIZANDO O ALGORITMO J48**. **Multi-Science Research**, Espirito Santo, jan./2019. Disponível em: http://msrreview.org/index.php/msr/article/viewFile/27/15. Acesso em: 5 out. 2019.

MUNDO EDUCAÇÃO. **A contribuição da Prova Brasil para o ensino público**. Disponível em: https://mundoeducacao.bol.uol.com.br/educacao/a-contribuicao-prova-brasil-para-ensino-publico.htm. Acesso em: 18 out. 2019.

NASCIMENTO, R. L. S. D; JUNIOR, G. G. D. C; FAGUNDES, R. A. D. A**. Mineração de Dados Educacionais: Um Estudo Sobre Indicadores da Educação em Bases de Dados do INEP**. 2016. Disponível em: <https://seer.ufrgs.br/renote/article/view/85989/49358>. Acesso em: abr. 2020.

PACHECO, Eliezer. **Os Institutos Federais: uma revolução na educação profissional e tecnológica. Secretaria da Educação Profissional e Tecnológica**, Ministério da Educação, [2008]. Disponivel em: <http://portal.mec.gov.br/index.php?option=com_docman&view=download&alias=3787-cartilha-eliezer-final&category_slug=marco-2010-p-df&Itemid=30192> Acesso em: mar. 2020.

PNE. **Meta 12 – Educação Superior**. Disponível em: http://simec.mec.gov.br/pde/grafico\_pne.php. Acesso em: 11 out. 2019.

PORTAL DO MEC. **Gestão da Educação Escolar**. Disponível em: <http://portal.mec.gov.br/seb/arquivos/pdf/profunc/06_gest_edu_esc.pdf>Acesso em: 4 abr. 2020.

RAMOS, J. L. C., SILVA, J., PRADO, L., GOMES, A., AND RODRIGUES, R. (2018). **Um Estudo Comparativo de Classificadores na Previsão da Evasão de Alunos em EAD.** In Anais do Simpósio Brasileiro de Informática na Educação (SBIE), volume 29, pages 1463–1472.

RIGOTTI, JIR; CERQUEIRA, CA. **As bases de dados do INEP e os indicadores educacionais: conceitos e aplicações.** Brasil, 2015. Disponível em: <http://www.abep.org.br/publicacoes/index.php/livros/article/viewFile/152/150>. Acessado em: 05 de abr. 2020.

ROMERO, C.; ROMERO, J.; VENTURA, S. **A survey on pre-processing educational data. Studies in Computational Intelligence**, v. 524, p. 29–64, 2014.

SANDER, Benno. **Política e Administração da Educação no Brasil:** momentos de uma história em construção. 2005

SANTOS, M. F.; AZEVEDO, C. S. **Data mining: descoberta de conhecimento em bases de dados**. São Paulo: FCA, 2005.

SILVA, É. R. D. A. **MINERAÇÃO DE DADOS EDUCACIONAIS GUIADO POR MAPAS DE CONHECIMENTO**. 2016. Disponível em: <http://www.uece.br/mpcomp/index.php/arquivos/doc_view/420-?tmpl=component&format=raw> . Acesso em: mar 2020.

SILVA, L. A. et al. **Ciência de Dados Educacionais: definições e convergências entre as áreas de pesquisa**. VI Congresso Brasileiro de Informática na Educação, Brasil, 2017. Disponível em: <https://www.br-ie.org/pub/index.php/wcbie/article/viewFile/7462/5258>. Acesso em: 27 set. 2020

SASSI, R. J. **Uma arquitetura Híbrida para Descoberta de Conhecimento em Base de Dados: Teoria de Rough Sets e Redes Neurais Aritificiais**. São Paulo, 2006. Disponível em: <https://www.teses.usp.br/teses/disponiveis/3/3142/tde-16032007-163930/publico/teseversaorevisada.pdf>. Acesso em: abr 2020.

SOUZA, Â. R. D. **As condições de democratização da gestão da escola pública brasileira**: Brasil, 2019. Disponível em: <http://www.scielo.br/pdf/ensaio/v27n103/1809-4465-ensaio-S0104-40362018002601470.pdf>. Acesso em: 9 mar. 2020.

TREVISAN, Leonardo; VELOZO, Elsa. **Gestão de competitividade e políticas públicas de formação de mão-de-obra: o caso Centro Paula Souza**, Brasil, 2007. Disponível em: http://www.scielo.br/pdf/rap/v41n5/a05v41n5.pdf. Acesso em: 20 fev. 2020.

VIEIRA, A. E. R; BUSSOLOTTI, Juliana Marcondes. **GESTÃO ESCOLAR: um estudo de caso sobre as Escolas Técnicas**, Brasil, 2018. Disponível em: <https://periodicos.unis.edu.br/index.php/interacao/article/view/167/152>. Acesso em: 9 mar. 2020.

YIN, R. K.; **Pesquisa Qualitativa do Início ao Fim**. 2.ed, Porto Alegre, 2016.

**Instruções gerais:**

O trabalho deverá ter no mínimo 20 e no máximo 30 páginas excluindo-se os elementos pré e pós-textuais, bem como as referências bibliográficas.

Os itens marcados como “**Obrigatório**” deverão necessariamente constar no projeto apresentado.

Os itens marcados como “**Opcional**” podem ou não constar no projeto a depender da linha de pesquisa do aluno (mas tendem a ser itens comumente requeridos em projetos da LP3), das orientações do professor responsável pela disciplina e do orientador. Também deverão ser observadas todas as recomendações e técnicas apresentadas na disciplina de Metodologia.