**PIPELINE** – METODOLOGIA DE ANÁLISE DE DADOS EDUCACIONAIS

ETAPAS DO PROCESSO:

**Entendimento de domínio:** Nesta etapa é realizado o entendimento do domínio ao qual os dados pertencem, em se tratando de dados educacionais, deve-se estender as particularidades do sistema educacional analisado, em relação ao Brasil, é importante compreender como são realizados os lavamentos dos dados e suas metodologias de tratamento utilizadas, as particularidade existentes entre os estados e suas regiões, funcionamento das políticas de educação e o contexto histórico da educação brasileira.

**Entendimento dos dados:** Nesta etapa preocupa-se em conhecer as características e particularidades dos dados disponíveis no conjunto de dados, atentando-se às escalas dos atributos numéricos e ás categorias existentes nos atributos categóricos, se estas categorias são ordinais ou nominais e por fim compreender o que estas escalas e categorias representam.

**Preparação dos dados:** Tratamento de dados nulos e faltantes; verificação de outliers; normalização; agrupamento dos dados de acordo com o objetivo desejado, os dados educacionais costumam separados por tipo de escola (pública, privada, estadual, federal, rural, urbana, etc.); integração de diferentes conjuntos de dados.

**Análise estatística e geração de informação:** Nesta etapa analisa-se os dados por meio de testes estatísticos como teste t e correlações, realiza-se a seleção dos atributos a serem estudados, gera-se gráficos de linha e barra e verifica-se resultados de média, mediana, desvio padrão, dentre outros.

**Avaliação:** Nesta etapa realiza-se a avaliação dos resultados, verifica-se possíveis padrões encontrados, se os atributos analisados apresentam alguma contribuição para a análise, se há alguma necessidade de se analisar outros atributos.

**Publicação de resultados:** Esta é a etapa final, após todas as análises concluídas, segue-se para a publicação dos resultados, que pode ser por meio de publicação de artigos, anuais, revistas científicas, relatórios a entidade requerente, publicações em jornais e afins.

**DIAGRAMA DE FLUXO:**

Preparação dos dados

Entendimento do domínio

Entendimento dos dados

Publicação

Avaliação

Análises estatísticas e GI