Alexandre de Souza Pereira – 20171102696.

O estudo presente nesta monografia, tem como objetivo principal fazer comparação entre as aplicações de modelos de aprendizado de máquina que melhor se encaixam com a base nos dados trabalhados. O aprendizado de máquina visa fazer o reconhecimento de padrões dentro da base de dados.  Porém antes de falar de fato dos modelos, temos pontos a ressaltar.

Iniciando o estudo de análise de dados, utilizaremos a linguagem de programação Python para o desenvolvimento de todo o projeto em questão. Há  bibliotecas que facilitam a manipulação dos dados, uma vez que a base de dados é muito grande e é necessário fazer algumas manipulações para o tratamento dos dados. Visto que o tratamento de dados é uma prática muito necessária, quando estamos desenvolvendo projetos é importante que consigamos identificar os outliers que podem expressar pontos como ausência de dados, dados com uma alta disparidade.

À medida que avançamos pelas partes mais iniciais do projeto, chegamos ao ponto que precisamos separar a base de dados para pôr em prática os devidos  treinamentos dos modelos. Para que não encontremos dificuldades quando obtemos os resultados perante os modelos, é necessário fazer a separação da base em algumas partes, como: base que serão submetidos os treinos,  para realização dos testes, para a validação dos modelos utilizados.

Ao passo que estamos fazendo a separação dos dados, precisa-se utilizar a normalização ou padronização, tendo em vista que alguns modelos fazem medição de distância entre os pontos, faz-se necessário utilizar desse facilitador matemático como citado anteriormente. Em resumo, é necessário utilizar o artifício, porém ficar atento para que os modelos não fiquem tendenciosos.