[**1 INTRODUÇÃO** 1](#_Toc150106581)

[1.1 Objetivo Geral 2](#_Toc150106582)

[1.2 Objetivos Específicos 2](#_Toc150106583)

[1.3 Justificativa 2](#_Toc150106584)

[2 REFERENCIAL TEÓRICO 4](#_Toc150106585)

[2.1 Descoberta de Conhecimento e a Mineração de Dados 4](#_Toc150106586)

[2.1.1 Seleção 5](#_Toc150106587)

[2.1.2 Pré-Processamento 6](#_Toc150106588)

[2.1.3 Transformação 6](#_Toc150106589)

[2.1.4 Mineração de Dados 7](#_Toc150106590)

[2.1.5 Interpretação e Avaliação 7](#_Toc150106591)

[2.2 Machine Learning 7](#_Toc150106592)

[2.2.1 Modelo Supervisionado 8](#_Toc150106593)

[2.2.2 Modelo Não Supervisionado 9](#_Toc150106594)

[2.3 Técnicas de Classificação 10](#_Toc150106595)

[2.3.1 Decision Tree 10](#_Toc150106596)

[2.3.2 Random Forest 11](#_Toc150106597)

[2.3.3 KNN 12](#_Toc150106598)

[2.3.4 Naive Bayes 13](#_Toc150106599)

[2.3.5 Logistic Regression 14](#_Toc150106600)

[2.4 Métricas de Avaliação 14](#_Toc150106601)

[2.4.1 Matriz de Confusão 15](#_Toc150106602)

[2.5 Satisfação em Serviços 17](#_Toc150106603)

[2.5.1 Satisfação de Serviços em Aeroportos 17](#_Toc150106604)

[2.6 Trabalhos Correlatos 18](#_Toc150106605)

[3 METODOLOGIA 23](#_Toc150106606)

[3.1 Descrição do Conjunto de Dados 23](#_Toc150106607)

[3.2 Atividades Desenvolvidas 23](#_Toc150106608)

[3.3 Ferramentas Utilizadas 25](#_Toc150106609)

[4. RESULTADOS E DISCUSSÕES 27](#_Toc150106610)

[4.1 Análise Descritiva 27](#_Toc150106611)

[4.2 Avaliação dos Modelos 30](#_Toc150106612)

[4.2.1 Sem Balanceamento de Classe 30](#_Toc150106613)

[4.2.2 Balanceamento por *Undersampling* 32](#_Toc150106614)

[4.2.3 Balanceamento por *Oversampling* 33](#_Toc150106615)

[4.2.4 *Oversampling* por *Resample* 34](#_Toc150106616)

[4.2.5 Melhores parâmetros obtidos pelo *Grid-Search* 35](#_Toc150106617)

[4.2.6 Avaliação do desempenho dos modelos 36](#_Toc150106618)

[4.3 Variáveis Mais Importantes para Classificação da Satisfação 37](#_Toc150106619)

[4.4 Avaliação do Modelos com Determinação Alternativa de Satisfeitos e Insatisfeitos 38](#_Toc150106620)

[4.4.1 Sem Balanceamento de Classe 38](#_Toc150106621)

[4.4.2 Balanceamento por *Undersampling* 39](#_Toc150106622)

[4.4.3 Balanceamento por *Oversampling* 41](#_Toc150106623)

[4.4.4 *Oversampling* por Resample 42](#_Toc150106624)

[4.4.5 Melhores Parâmetros Obtidos pelo *Grid-Search* 43](#_Toc150106625)

[4.4.6 Avaliação de Desempenho dos Modelos 44](#_Toc150106626)

[4.5 Análise de Usuários com Satisfação Geral Intermediária 45](#_Toc150106627)

[4.5.1 Análise Exploratória dos Usuários com Satisfação Intermediária 45](#_Toc150106628)

[4.5.2 Avaliação dos Modelos dos Usuários com Satisfação Intermediária 48](#_Toc150106629)

[**5 Conclusão** 50](#_Toc150106630)

[REFERÊNCIAS 52](#_Toc150106631)

[APÊNCIDE A 57](#_Toc150106632)

[APÊNDICE B 61](#_Toc150106633)