W celu wykonania badań eksploracyjnych (EDA, ang. exploratory data analysis) dla zestawu danych wine.data został przygotowany program w języku R. W pierwszej kolejności zostały wczytane potrzebne biblioteki oraz załadowane dane wine.data. Następnie program wyświetlił pierwsze 15 linii zestawu danych oraz wyznaczył jego wymiary i strukturę.

Zestaw danych wine.data składa się z 178 próbek opisujących 3 rodzaje wina, co wynika z wyświetlonego wcześniej rozkładu liczby próbek dla każdej klasy. Struktura zestawu danych obejmuje 13 cech objaśniających oraz 1 cechę decyzyjną (klasy wina).

Następnie program obliczył standardowe odchylenie kwadratowe, skośność oraz kurtozę dla każdej cechy liczbowej. Zostały one wyświetlone jako wektory dla każdej cechy.

Następnie zostały wyznaczone cecha decyzyjna (klasy wina) oraz cechy objaśniające, które zostały zapisane do odpowiednich zmiennych.

Program przygotował także wizualizację danych, tworząc histogramy dla cech typu numerycznego oraz wykresy słupkowe dla typów jakościowych. Został również zbudowany diagram analizy korelacji dla wszystkich cech objaśniających. Zostały także przygotowane wykresy pudełkowe dla atrybutów typu liczbowego oraz wykresy rozrzutu dla cech typu numerycznego w zależności od klas. Program sporządził również wykresy rozkładu empirycznego (gęstości) dla cech typów liczbowych w zależności od klas.

Następnie program podzielił zbiór danych na uczący i testowy, z zastosowaniem 75% danych w zbiorze uczącym. Następnie zostały wykonane klasyfikacje za pomocą algorytmów następujących algorytmów: naiwny Bayes, K najbliższych sąsiadów, maszyny wektorów nośnych, drzewa klasyfikacyjnego oraz lasu losowego. Skuteczność każdego z tych algorytmów została oszacowana i porównana na podstawie miar oceny klasyfikacji: Accuracy oraz Kappa Cohena.

Następnie program wybrał najskuteczniejszy algorytm i zastosował go do klasyfikacji zbioru testowego. Ostatecznie została obliczona macierz błędów oraz podstawowe miary jakości dla wybranego algorytmu.

Po przeprowadzeniu wszystkich powyższych kroków, uzyskano pełen obraz zestawu danych wine.data oraz zostały zastosowane różne metody klasyfikacji, z których najskuteczniejszy został wybrany i zastosowany do klasyfikacji zbioru testowego. Rezultaty tych działań pozwolą na lepsze zrozumienie i interpretację danych oraz na wybór odpowiedniego algorytmu do ich analizy.