Contains the data employed in the manuscript titled: “Scalable Flight Cancellation Prediction with Ensemble Distributed KNN and Feature Selection.” the research data has been collected through the FlightRadar24 website using a web crawler and targets the years 2021 to 2023. Each data record has been described through 36 indicators, which fall into three general classes: weather (7 indicators), flight features (19 indicators), and airport features (10 indicators). The collected samples have been divided into four classes based on the flight cancellation/delay status: 1- OnTime, 2- Late, 3- Very Late, and 4- Cancelled.