ct - light delay prediction is fundamental to establish the more efficient airline business. The development of accurate prediction models for flight delays became cumbersome due to the complexity of airtransportation system, the number of methods for prediction, and the deluge of flight data. In this context, this paper presents a thorough literature review of approaches used to build flight delay prediction model. Airlines delays make immense loss for business field as well as in budget loss for a country. Flight delays hurt airlines, airports, and passengers. We are proposing machine learning algorithms like Linear regression Techniques. The aim of this research work is to predict Flight Delay, Which is highest economy producing field for many countries and among many transportation this one is fastest and comfort, so to identify and reduce flight delays, can dramatically reduce the flight delays to saves huge amount of turnovers, using machine-learningalgorithms. Flight delays could always be annoying, especially in the case when the period of delay was so long that there was even a danger to miss the next flight. However, if there was a way to predict whether there would be a delay or even better – how long the delay could be, then people could make earlier preparation to reschedule following flights in an earlier manner.Index Terms - Data prediction, Machine Learning

.LITERATURE SURVEY

A.LOGISTIC REGRESSION MODELNo data mining projects could be finished without thoroughly understand the data first. So, in order to better understand data, we start our project by exploring the data first. We found the original dataset includes 28 attributes/columns and while most of the data were in float format, some of them were object types). In addition, as shown, there were also many null values in the original datasets. So, we need to first clean the columns with null values and change data types of objects into suitable types (mostly integers) for the convenience of machine learning.

B.LINEAR REGRESSION MODELWe renamed the original data column names and validated the nulls, however with a little different approach. We first plotted a density plot for chosen attributes. After plotting the density plot with columns with “nan” values, we found none of the columns strictly follows normal distribution, and most of them were largely skewed and concentrated to only few values. Replacing methods, we tried included applying fillna () method to replace “nan” and replacing missing “nan” values with the mean of corresponding columns. However, none of the methods enable us to develop model with desirable results. So instead of replacing “nan” with normal distribution, we decided to use merely replace “nan” with extreme values that without the original data range.

C.INITIAL DATA EXPLORINGAfter data cleaning we start the first process of exploring our data if there were any patterns within the independent variables. The above graph shows the no of delays airline wise. On the left side you can see there is a false value, which means instances when an airline has not been delayed