

RESEARCH PAPER

ABSTRACT

Topic : Diabetes Prediction using Machine Learning with PySpark MLlib.

PySpark is a collaboration between Apache Spark and Python. PySpark is a tool used in Big Data Analytics. It provides a wide range of libraries and is majorly used for Machine Learning and Real-Time Streaming Analytics. In this paper we discuss about how to make more accurate predictions for diabetes prediction than the existing prediction methods. The existing methods are not possible to be performed by the general public as a programming knowledge is required to perform all the operations. This paper also speaks about the ways we can reduce the complexity of the algorithms that need to be performed to get an accurate prediction.