Kepler Exoplanet Search Results using Machine Learning Classification Model

Sasmita Kumari Nayak[1], Nikita padhan[2], Atasi pradhan[3] and Mohammed Siddique[3*]
[1]Department of CSE, Centurion University of Technology and Management, Odisha, India
[2,3]School of Applied Sciences, Centurion University of Technology and Management, Odisha
[1]Department of Mathematics, Centurion University of Technology and Management, Odisha, India
Corresponding author Mail id: siddique1807@gmail.com, siddique@cutm.ac.in

Abstract
The Kepler exoplanet mission is specially organised for searching of the Milky Way galaxy to discover many earth-size and tiny planets near or in the habitable zone. Exoplanet means orbiting of planet around the other stars beyond our solar system. The dataset contains 9564 samples and 50 columns which is collected from kaggle website. The dataset, target variable is KOI-disposition which contains confirmed, false positive and candidate. Out of 9564 samples we found 5000 samples are false positive, each confirmed and candidate are 2282. In the Milky Way galaxy many stars and planets are there, but we have considered some of them. We have used machine learning algorithms like decision tree, random forest, KNN classification and Naive Bayes classification on stars and planets for searching of exoplanets beyond our stellar atmosphere.
Keywords: Exoplanets; Kepler data; Solar System, Koi, Decission Tree, Random Forest, KNN.

1. Introduction

The Kepler space observatory is a NASA build satellite that was launched in 2009. The telescope is dedicated to searching for exoplanets in star systems beside our own , with the ultimate goal of possibly finding other habitable planets besides our own. The mission results provided data on a wide range of planet and planetary systems orbiting both single and multiple stars of differing sizes , temperatures and ages.

The original mission ended in 2013 due to some technical failures, but the telescope has still been functional since 2014 on a “K2” extended mission. The telescope is still alive and continuously collects new data on its extended mission. The main objective of our project is searching of exoplanets using kepler data like koi_slogg , koi_impact , koi_depth , koi_period , koi_duration , koi_impact and koi_score to get the target variable i.e. koi_disposition which contain false positive , confirmed and candidate. Till the effects of these biases are correctly determined and the full range of orbital periods are considered, estimations of parent distributions come with large uncertainties.

2. Implementation of the Model

Model implementation is a most important part for a machine learning project. In this part we have to chose the columns for feature selection then remove all other unnecessary columns after