Readme

1) Run the command: python create_collections.py
   a) This converts the whole data set into a pickle file named ‘Patients’. This conversion may take 15 mins as the data set is large.
2) Run the command: python prediction.py
   a) This uses the Patients pickle file and creates another pickle ‘NewFeatures’ file which has the selected features data in a pandas dataframe.
   b) This selected features are used to predict the onset of sepsis disease using Logistic Regression, Extra Trees, Support Vector Machine, Gradient Boost, KNN classifiers before 4 & 6 hours.
   c) The run time of this command may take 10-15 mins since all the prediction algorithms run for both 4 & 6 hours

Dependencies:

Install the following packages before running the above mentioned commands.

1) Numpy
2) Scikit-learn
3) Pandas
4) SciPy
5) matplotlib
6) imbalanced-learn

Note:

- Make sure that all the files and folders must be in the same directory.
- Details of methodology followed and results will be included in the presentation.