**Capstone Project**

**Predicting Liver disease from data**

**Machine Learning Engineer Nanodegree**

**Shaik Abdul Haleem**

09/03/2019

***Definition***

**Project Overview**

In India, delayed diagnosis of diseases is a fundamental problem due to a shortage of medical

professionals. A typical scenario, prevalent mostly in rural and somewhat in urban areas is:

1. A patient going to a doctor with certain symptoms.

2. The doctor recommending certain tests like blood test, urine test etc depending on the symptoms.

3. The patient taking the aforementioned tests in an analysis lab.

4. The patient taking the reports back to the reports back to the hospital, where they are examined

and the disease is identified.

The aim of this project is to somewhat reduce the time delay caused due to the unnecessary back

and forth shuttling between the hospital and the pathology lab. Historically, work has been done in

identifying the onset of diseases like heart disease, Parkinson's from various features, for example in

this paper <https://link.springer.com/chapter/10.1007/978-3-319-11933-5_17.In> this case, a machine

learning algorithm will be trained to predict a liver disease in patients.

Problem Statement

The problem statement is formally defined as:

‘Given a dataset containing various attributes of 583 Indian patients, use the features available in

the dataset and define a supervised classification algorithm which can identify whether a person is

suffering from liver disease or not.’

The dataset for this problem is the ILPD (Indian Liver Patient Dataset) taken from the UCI Machine

Learning Repository . Number of instances are 583. It is a multivariate data set, contain 10 variables

that are age, gender, total Bilirubin, direct Bilirubin, total proteins, albumin, A/G ratio, SGPT, SGOT

and Alkphos. All values are real integers. This data set contains 416 liver patient records and 167

non- liver patient records.The data set was collected from north east of Andhra Pradesh, India. This

data set contains 441 male patient records and 142 female patient records. Any patient whose age

exceeded 89 is listed as being of age "90".

Relevant sources: Bendi Venkata Ramana, Prof. M. S. Prasad Babu and Prof. N. B. Venkateswarlu,

â€œA Critical Comparative Study of Liver Patients from USA and INDIA: An Exploratory Analysisâ€•,

International Journal of Computer Science Issues, ISSN :1694-0784, May 2012.