Goal

The minimum goal of this project is to perform the **prediction** of the **workload** of a **web server** by training an LSTM neural network on the web server **log**s.
The optional goals are to perform anomaly detection and diagnosis.

Domain

Log analysis.
My familiarity with the domain comes from my previous experience as integration software engineer for which I have been exposed to distributed systems log analysis.

Data

The dataset for this project is seven month's worth of all HTTP requests to the University of Saskatchewan's WWW server. (http://ita.ee.lbl.gov/html/contrib/Sask-HTTP.html)
The logs are an ASCII file with one line per request, with the following columns.

Variable Name	Type	Description
remotehost	String	Remote hostname (or IP number if DNS hostname is not available).
rfc931	String	Remote logname of the user.
date	Datetime	Date and time of the request
request	String	The request from the browser or client.
status	String	The HTTP status code the

		server sent back to the client
bytes	integer	The number of bytes (Content-Length) transferred to the client.

Tools

The LSTM neural network would be implemented in Keras.

Known Unknowns

The dataset is not very big, maybe the anomaly detection would not be possible.