Team02
Machine Learning Dataset Parsing Tool

Group Members
Boyuan Ma - Team Leader
Xinyan Li
Hao Liu
Xinjie Pang
Marios Igkiempor - scymi1

Supervisor
Chris Roadknight

December 2018

Contents

1 Description of the Problem

Our team's project is centered around a system that classifies datasets by the best type of machine learning approach to take in order to best analyse the data. It is proposed that the system will be able to parse a dataset, analyse its features and propose a type of learning (supervised, semi-supervised and unsupervised) that can best model the dataset based on the analysed features.

The client intends to use the project as both a prefilter for machine learning and as a teaching aid, and so the program should be able to provide useful information as to how the machine learning approach was derived.

Datasets that the program will analyse will be provided by the client. The possibility of analysing new datasets will be a target that the team will strive for, however it is not essential for the core functionality required by the client.

In essense, the project will take a dataset and tell the user the most optimal machine learning strategy with which to analyse the data.

2 Background Information and Relevant Research

3 Requirements Specification

4 Initial Design Ideas

5 Key Implementation Decisions

5.1 Database Design

6 Problems Encountered

7 Time Planning and Project Management