

A report on Scikit Learn

An assignment submitted in the partial fulfillment of the
requirement for the degree of

**MASTER OF TECHNOLOGY in COMPUTER
SCIENCE AND ENGINEERING**

Submitted to
Dr. Annappa B

Submitted by
Hradesh Patel
172CS012

Srikant Singh
172CS022

Harish Kumar
172CS011



DEPARTMENT OF COMPUTER SCIENCE
NATIONAL INSTITUTE OF TECHNOLOGY
KARNATAKA
SURATHKAL, MANGALORE 575025

Contents

1	ABSTRACT	3
2	INTRODUCTION	4

List of Figures

1 ABSTRACT

Scikit-learn has emerged as one of the most popular toolkits for machine learning, and is widely used in industry and academia. It is a Python module integrating a wide range of state-of-the-art machine learning algorithms for medium-scale supervised and unsupervised problems. This package focuses on bringing machine learning to non-specialists using a general-purpose high-level language. Emphasis is put on ease of use, performance, documentation, and API consistency. It has minimal dependencies and is distributed under the simplified BSD license, encouraging its use in both academic and commercial setting. It features various classification, regression and clustering algorithms including support vector machines, random forests, gradient boosting, k-means and DBSCAN, and is designed to interoperate with the Python numerical and scientific libraries NumPy and SciPy.

Scikit-learn exposes a wide variety of machine learning algorithms, both supervised and unsupervised, using a consistent, task-oriented interface, thus enabling easy comparison of methods for a given application. Since it relies on the scientific Python ecosystem, it can easily be integrated into applications outside the traditional range of statistical data analysis. Importantly, the algorithms are implemented in a high-level language.

2 INTRODUCTION