CITY College

University of York Europe Campus

Computer Science Department

UNDERGRADUATE INDIVIDUAL PROJECT

Algorithmic Showdown: Unveiling the Best in Classification

This report is submitted in partial fulfillment of the requirement for the degree of Bachelors in Computer Science with Honours by

Aleksandar Radevski

May 2024

Approved

Dr. Ourania Mangira

_____

_____

Contents

1. Introduction

Nowadays, classification algorithms are an indispensable part of machine learning. They help solve complex real-world problems. From predicting customer churn to diagnosing or even predicting the possibility of the existence of a disease, these types of algorithms play a major role in the machine learning industry. But in order for them to be at a high level, the datasets used for training and evaluation must have precisely defined characteristics and a clearly defined evaluation value.

A base or even a key point in evaluating and comparing classification algorithms is the reading of datasets themselves. Many of these datasets are either of limited scope or have interference that drastically affects the final results. For the purpose of this coursework, an artificial (synthetic) dataset was created that is easy to read and at the same time satisfies the requirements for easy and at the same time conscious comparison of different classification algorithms.

This dissertation focuses on two main goals: first, to see, compare and describe the efficiency, accuracy, and last but not least the ease of writing of different classification algorithms; and second, to demonstrate the importance of data quality and design in achieving satisfactory results. Specifically, this paper examines the performance of algorithms on balanced versus unbalanced data and complete versus incomplete data sets.