A Mini Project Report on

“Mushroom Classification Using Genetic Algorithm”

Submitted By

Shubham Kokane (BECOMPA-35)
Shreyas Soni (BECOMPA-68)
Piyush Wadi (BECOMPA-72)

(Final Year Computer Engineering)

Guided By

Dr. Archana Chaugule

Department of Computer Engineering
Pimpri Chinchwad College of Engineering and Research, Pune 412101 [2020-21]

PIMPRI CHINCHWAD EDUCATION TRUST'S

Pimpri Chinchwad College of Engineering and Research, Pune 412101

C E R T I F I C A T E

This is to certify that *(Shubham Kokane, Shreyas Soni, Piyush Wadi) has* successfully completed the project entitled "Mushroom Classification Using Genetic Algorithm" in the fulfillment of B. E. (Computer Engineering) LP-IV and this work has been carried out in my presence.

Date :

Place :

Dr. Archana Chaugule
Guide and HOD
(Dept. of Comp. Engg.)
Pimpri Chinchwad College of Engineering
and Research, Pune 412101

External

Prof. Dr. Tiwari H.U.
Principal,
Pimpri Chinchwad College of Engineering
and Research, Pune 412101

ACKNOWLEDGEMENT

This is a great pleasure & immense satisfaction to express my deepest sense of gratitude & thanks to everyone who has directly or indirectly helped me in completing my project work successfully.I express my gratitude towards Project guide (Dr. Archana Chaugule) , Head of Department of Computer Engineering, Pimpri Chinchwad college of Engineering and Research, Pune 412101 who guided & encouraged me in completing the project work in scheduled time. I would like to thank our Principal (Prof. Dr. Tiwari H.U), for allowing us to pursue my project in this institute.No words are sufficient to express my gratitude to our parents for their unwavering encouragement. We also thank all friends for being a constant source of my support.

Shubham Kokane
Shreyas Soni
Piyush Wadi

ABSTRACT

Genetic Algorithm (GA) is a search-based optimization technique based on the principles of **Genetics and Natural Selection**. It is frequently used to find optimal or near-optimal solutions to difficult problems which otherwise would take a lifetime to solve. We are using Genetic Algorithm to classify mushrooms based on its characteristics. The genetic algorithm depends on selection criteria, crossover, and mutation operators. . In the implementation, cities are taken as genes, string generated using these characters is called a chromosome, while a fitness score which is equal to the path length of all the cities mentioned, is used to target a population.Fitness Score is defined as the length of the path described by the gene. Lesser the path length fitter is the gene. The fittest of all the genes in the gene pool survive the population test and move to the next iteration.

In computer science and operations research, a genetic algorithm (GA) is a metaheuristic inspired by the process of natural selection that belongs to the larger class of evolutionary algorithms (EA). Genetic algorithms are commonly used to generate high-quality solutions to optimization and search problems by relying on biologically inspired operators such as mutation, crossover and selection.John Holland introduced genetic algorithms in 1960 based on the concept of Darwin's theory of evolution; his student David E. Goldberg further extended GA in 1989

INDEX

List Of Figures

1. INTRODUCTION

Genetic algorithms are based on an analogy with genetic structure and behavior of chromosomes of the population. Following is the foundation of GAs based on this analogy

1. Individual in population compete for resources and mate
2. Those individuals who are successful (fittest) then mate to create more offspring than others
3. Genes from “fittest” parents propagate throughout the generation, that is sometimes parents create offspring which are better than either parent.
4. Thus each successive generation is more suited for their environment.

The genetic algorithm determines which subset of features has the best fitness as the input for mushroom type classification.

2. MOTIVATION

Nature has always been a great source of inspiration to all mankind. Genetic Algorithms (GAs) are search based algorithms based on the concepts of natural selection and genetics. GAs are a subset of a much larger branch of computation known as **Evolutionary Computation**.

GAs were developed by John Holland and his students and colleagues at the University of Michigan, most notably David E. Goldberg and has since been tried on various optimization problems with a high degree of success.

In GAs, we have a **pool or a population of possible solutions** to the given problem. These solutions then undergo recombination and mutation (like in natural genetics), producing new children, and the process is repeated over various generations. Each individual (or candidate solution) is assigned a fitness value (based on its objective function value) and the fitter individuals are given a higher chance to mate and yield more “fitter” individuals. This is in line with the Darwinian Theory of “Survival of the Fittest”.

In this way we keep “evolving” better individuals or solutions over generations, till we reach a stopping criterion.

Genetic Algorithms are sufficiently randomized in nature, but they perform much better than random local search (in which we just try various random solutions, keeping track of the best so far), as they exploit historical information as well.

Thus in this case, the genetic algorithm determines which subset of features has the best fitness as the input for mushroom type classification.

3. PROBLEM STATEMENT

Apply the Genetic Algorithm for classifying mushrooms. There are various types of Mushrooms, some are edible and others are not. By applying a genetic algorithm, we try to determine whether a specific mushroom is edible or not on a given dataset.

4. OBJECTIVES

Objectives of the project are:

- Use genetic algorithm to distinguish edible and non-edible mushrooms
- Show output
- Analyze the improvements in the solution

5. SOFTWARE AND HARDWARE REQUIREMENTS

5.1 SOFTWARE REQUIREMENTS:

1. Python
2. PIL
3. Numpy
4. Matplotlib

5.2 HARDWARE REQUIREMENTS:

1. Hard Disk: minimum of 1GB of available hard disk space.
2. Memory: recommended 4GB
3. Operating System: *Windows / Ubuntu.*

5. Tool: Introduction

1.Python:

Python is a widely used general-purpose, high level programming language. It was created by Guido van Rossum in 1991 and further developed by the Python Software Foundation. It was designed with an emphasis on code readability, and its syntax allows programmers to express their concepts in fewer lines of code.

Python is a programming language that lets you work quickly and integrate systems more efficiently.

There are two major Python versions: **Python 2 and Python 3**.

2.Numpy

NumPy is a Python package. It stands for 'Numerical Python'. It is a library consisting of multidimensional array objects and a collection of routines for processing of arrays.

Numeric, the ancestor of NumPy, was developed by Jim Hugunin. Another package Numarray was also developed, having some additional functionalities. In 2005, Travis Oliphant created the NumPy package by incorporating the features of Numarray into the Numeric package. There are many contributors to this open source project.

3.Pandas

Pandas is an open-source, BSD-licensed Python library providing high-performance, easy-to-use data structures and data analysis tools for the Python programming language. Python with Pandas is used in a wide range of fields including academic and commercial domains including finance, economics, Statistics, analytics, etc. In this tutorial, we will learn the various features of Python Pandas and how to use them in practice.

6.Matplotlib

Matplotlib is one of the most popular Python packages used for data visualization. It is a cross-platform library for making 2D plots from data in arrays. Matplotlib is written in Python and makes use of NumPy, the numerical mathematics extension of Python. It provides an object-oriented API that helps in embedding plots in applications using Python GUI toolkits such as PyQt, WxPython Tkinter. It can be used in Python and IPython shells, Jupyter notebook and web application servers also.

Matplotlib has a procedural interface named Pylab, which is designed to resemble MATLAB, a proprietary programming language developed by MathWorks. Matplotlib along with NumPy can be considered as the open source equivalent of MATLAB.

7. STEPS OF IMPLEMENTATION WITH OUTPUT

- **7.1 Steps of implementation :**
    1. Load dataset
    2. Split into training and testing dataset.
    3. Implement Genetic Algorithm
    4. Calculate accuracy.

- **7.2 Output Screenshots**

    1. **Accuracy increment during iterations**

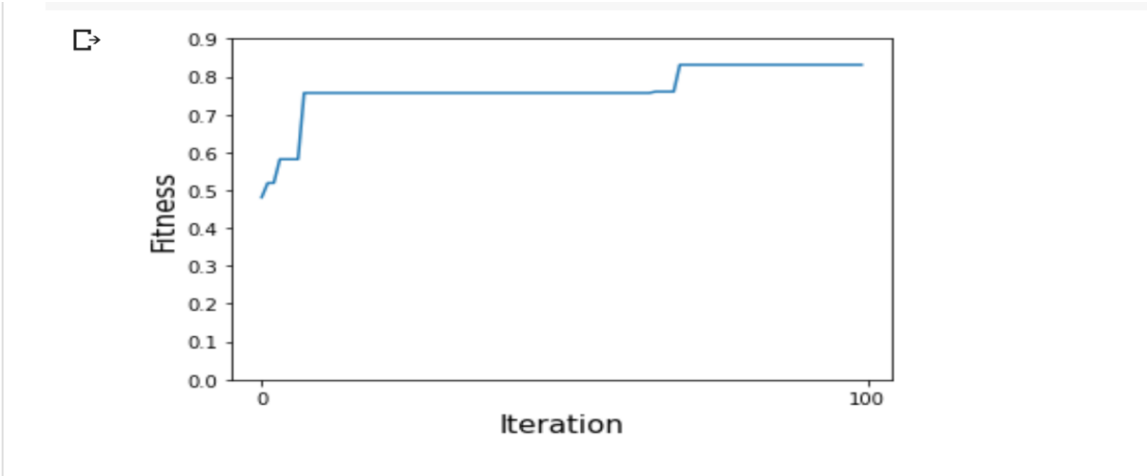

Fig 7.2.1 Accuracy during 100 Generations

8. CONCLUSION

Hence, using genetic algorithms we can manage data sets with many features and also it performs better than traditional feature selection techniques since it doesn't require any specific knowledge about the problem which can be used further for parallelism in computer clusters.